DOCTORAL THESIS

Training-related risk factors and genetic polymorphisms associated with overuse injuries in recreational runners.

Kozlovskaia, Mariia

Award date:
2019

BOND UNIVERSITY

Training-related risk factors and genetic polymorphisms associated with overuse injuries in recreational runners

Mariia Kozlovskaia

Faculty of Health Sciences and Medicine

Associate Professor Kevin Ashton

A thesis submitted in total fulfilment of the requirements of the degree of *Doctor of Philosophy (PhD)*

July 2019

This research was supported by the Collaborative Research Network for Advancing Exercise & Sports Science Scholarship

Abstract

Recreational running is a type of physical activity with proven health benefits, including improvements in aerobic fitness and cardiovascular function. Recreational running is reported as one of the most popular physical activities among Australians, with 7.4 of the population participating. As with any physical activity, there are health-related risks, with musculoskeletal injuries the most commonly observed injury in runners, specifically, Achilles tendon injuries and bone stress injuries. Running-related injuries are complex conditions, and their development depends on the interaction of multiple factors. These include extrinsic factors, such as training, lifestyle habits, and intrinsic factors, such as sex, age and genetic predisposition. Musculoskeletal injuries are recurrent in nature, preventing runners from their regular physical activity and therefore affect their health. An increased understanding of factors that contribute to the development of the injuries may, therefore, assist with injury prevention strategies, reduce injury rates and potentially encourage more people to exercise regularly. This study aimed to investigate the health, training and lifestyle habits of Australian recreational runners, identifying the most common running-related injuries and associated risk factors, and explore genetic variants associated with Achilles tendon injuries and bone stress injuries using a genome-wide association study (GWAS) approach.

The project utilised an online survey platform and versatile promotional approaches in order to recruit recreational runners across Australia. A 25-month recruitment campaign resulted in a dataset of 4,720 completed survey responses. This study demonstrated that our cohort of recruited Australian recreational runners met recommended physical activity guidelines; and health risk factors such as smoking, overweight and hypertension were not typical. Moreover, the commencement of a running program was associated with significant weight loss. In addition, this study identified significant differences in running habits of male and female runners, with men were more likely to run more than six sessions per week and at a faster race pace than female recreational runners. In relation to injuries, an Achilles tendon injury was the most commonly reported injury, with its occurrence associated with being male, older in age, and faster race pace and stretching. The second most commonly reported injury was a bone stress injury, and its occurrence was associated with younger age, obesity, and longer weekly running distance and stretching.

Whilst a large-scale survey (4,720 participants) was achieved (Chapters 2 and 3), and despite numerous adjustments and targeted recruitment strategies, our recruitment into the

GWAS were smaller than expected. The data and findings presented in Chapters 4 and 5 are from this initial explorative study of genetic association with Achilles tendon injuries and bone stress injuries. Challenges of recruitment for large-scale genetic studies are also discussed throughout this thesis.

The GWAS component of the presented study comprises 1,099 analysed samples achieving only 23.1 and 95.8 of our case-control recruitment targets, respectively. The genetic data analyses were of an explorative nature as the sample size was not sufficient for reaching the significance levels required for the GWAS. The case-control analysis of the genetic variants associated with Achilles tendon injuries identified several putative genes (*TCF7L1*, *DOCK4* and *TLE1*), which were linked to the Wnt signalling pathway. The study failed to replicate the results of previously reported genetic associations with Achilles tendon injuries, except two results of polymorphisms which were a putative replication (rs1110495 in the intragenic region (6:51914974) and rs12722 in *COL5A1* gene). These results should be interpreted with caution due to the imputed nature of the data. The second case-control analysis of genetic association with bone stress injuries identified only one genetic variant in the gene *TMEM135*, whose protein function may contribute to the development of these injuries. Attempts to replicate previously shown genetic associations did not confirm those findings, except one putative replication for rs2051748 in *CALCR* gene. However, identified associations have to be taken with caution due to the small sample size of the study and risk of false-positive results.

In conclusion, this study demonstrated that Australian recreational runners had a body mass index within the healthy weight range, seemed to be meeting the recommended physical activity guidelines and were non-smokers. The identified sex differences in training habits may contribute to the promotion of recreational running depending on biological sex. The described factors associated with running-related injuries, specifically Achilles tendon injuries and bone stress injuries contributed to the body of knowledge about the development of running-related injuries and potentially would advance injury preventive strategies. The analyses of the genetic data provided new information about potentially important genetic markers and pathways, which may contribute to the development of running-related injuries and further our understanding of the pathophysiology behind these injuries. The collected phenotypic and genomic data may also contribute to the future research of these injuries, data meta-analysis and a better understanding of the genetics of multifactorial conditions.

Keywords: Physical activity, Running, Injury, GWAS, genetics, Achilles tendon, bone stress injury, Weight

Declaration by candidate

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy (PhD). This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made.

Mariia Kozlovskaia

July 2019

Research outputs arising from the thesis

Publications:

1. Kozlovskaia, M., Vlahovich, N., Ashton, K.J. and Hughes, D.C., 2017. Biomedical Risk Factors of Achilles Tendinopathy in Physically Active People: A Systematic Review. Sports Medicine-Open, 3(1), p.20.
2. Kozlovskaia M., Vlahovich N., Rathbone E., Manzanero S., Keogh J., Hughes D.C., 2017. A profile of health, lifestyle and training habits of 4720 Australian recreational runners – the case for promoting recreational running for health benefits. Health Promotion Journal of Australia.
3. Manzanero S., Kozlovskaia M., Vlahovich N., Hughes D.C., 2018. Recruitment and Participation of Recreational Runners in a Large Epidemiological and Genetic Research Study: Retrospective Data Analysis. JMIR Research Protocols.

Oral and poster presentations:

1. 2015 - ASICS Sports Medicine Australia Conference, Gold Coast, QLD. Poster presentation: ‘Biomedical risk factors of Achilles Tendinopathy: a systematic review.’
2. 2015 - PhD European Molecular Biology Laboratory (EMBL) Australian Conference, Perth, WA. Poster presentation: ‘Genetics of exercise-induce injuries in tendon and bone. The “AIS Injury Study”.’
3. 2016 - HSM Conference, Bond University, Gold Coast, QLD. Oral presentation: ‘Epidemiological characteristics of Australian recreational runners. The “AIS Injury Study”.’
4. 2017 - CRN-AESS Wrap up Symposium, Gold Coast, QLD. Oral presentation: ‘Genetics of exercise induced injuries in tendon and bone – first results.’
5. 2017 - New Investigator Forum, Canberra, ACT. Poster presentation: ‘A health and training profile of 4720 Australian recreational runners.’
6. 2017 - CHARM Conference, Canberra, ACT. Poster Presentation: ‘Risk factors of running-related injuries in Australian recreational runners.’

Ethics declaration

The research associated with this thesis received ethics approval from the Bond University Human Research Ethics Committee. Ethics application number RO1688B (Appendices 3 and 4).

Acknowledgements

I would like to sincerely thank my supervisors Associate Professor Kevin Ashton, Dr David Hughes, Dr Nicole Vlahovich and Dr Paul Leo for their expert advice, assistance and support during the project. I would like to acknowledge the Collaboration Research Network for Advancing Exercise & Sports Science (CRN AESS) for providing me with the PhD scholarship and the Health Science and Medicine Faculty of Bond University for the support during my enrolment. I would like to thank the Department of Sports Medicine at the Australian Institute of Sport for the opportunity to be a part of the department and extensive support during my PhD project and my integration into Australian community. Especially I would like to thank Dr Nicole Vlahovich for her patient guidance, mentorship and enthusiastic encouragements during this project.

I would like to thank Professor Matt Brown, Associate Professor Justin Keogh, Ms Evelyne Rathbone, Ms Rebecca Grealy, and the extended research team of the CRN AESS for sharing their expertise and scientific advice that substantially contributed to this project's outcomes.

I would like to thank all Australian recreational runners who volunteered for this project. Without their participation this study would not have been accomplished.

I would like to thank my husband, family and close friends for their tremendous support, encouragement and understanding during this PhD project. To my father and mother, thank you for always believing in me and supporting all my endeavours.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACE	Angiotensin converting enzyme
ACL	Anterior cruciate ligament
ADAM12	ADAM metallopeptidase domain 12 gene
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
AGT	Angiotensin
AGTR1	Angiotensin II receptor
AIS	Australian Institute of Sport
AT	Achilles tendon
BACH2	BTB domain and CNC homolog 2 gene
BDNF	Brain-derived neurotrophic factor
BMD	Bone mineral density
BMI	Body mass index
BMP2	Bone morphogenetic protein 2 gene
BS	Bone stress
CALCR	Calcitonin receptor gene
CASP	Caspase
CFTR	Cystic fibrosis transmembrane conductance regulator gene
CLIC4	Chloride intracellular channel 4
CNV	Copy number variant
COL11A1	α chain of the type XI collagen gene
COL12A1	α chain of the type XII collagen gene
COL14A1	α chain of the type XIV collagen gene
COL27A1	α chain of the type XXVII collagen gene
COL5A1	α chain of the type V collagen gene
CSMD	CUB and Sushi multiple domain
CSMD2	CUB and Sushi multiple domains 2 gene
CVD	Cardiovascular disease
DDX46	DEAD-box helicase 46 gene
DNA	Deoxyribonucleic acid

DOCK4	Dedicator of cytokinesis 4 gene
ELN	Elastin gene
ESR1	Estrogen receptor 1
FBN2	Fibrillin 2 gene
GDF5	Growth/differentiation factor-5
GSTA	Glutathione S-transferase
GSTA2	Glutathione S-transferase alpha 2 gene
GWAS	Genome-wide association study
HDL-C	High density lipoprotein cholesterol
HWE	Hardy-Weinberg equilibrium
IBS	Identity by state
IL1B	Interleukin 1β gene
IL1RN	IL-1β receptor antagonist gene
IL-1β	Interleukin-1 β
IL-4	Interleukin-4
IL-6	Interleukin-6
IQR	Interquartile range
ITB	Iliotibial band
LD	Linkage disequilibrium
LRP5	Low-density lipoprotein receptor 5 gene
LYPD6B	LY6/PLAUR domain containing 6B gene
MACROD1	MACRO domain containing 1 gene
MAF	Minor allele frequency
MIR1270	MicroRNA 1270 gene
MIR608	MicroRNA 608 gene
MMP	Matrix metalloproteinase
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MTSS	Medial tibial stress syndrome
NF-κB	Nuclear factor κB
NLGN1	Neuroligin 1 gene

NOS Nitric oxide synthase

NRXN3 Neurexin 3 gene

NTRK2 Neurotrophic receptor tyrosine kinase 2 gene

OPG Osteoprotegerin

OR Odds ratio

P2X7 Purinergic receptor gene

PDZRN4 PD domain containing ring finger 4 gene

PRIM2 Primase subunit 2 gene

QC Quality control

RANK Receptor activator of nuclear factor-KB gene

RANKL Receptor activator of nuclear factor-KB ligand gene

RARS Arginyl-tRNA synthetase gene

RED-S Relative Energy Deficiency in Sport

RI Running Injury

RR Relative risk

RUNX3 Runt-related transcription factor 3 gene

SLC35F3 Solute carrier family 35, member F3 gene

SMARCD1 SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member

SNP Single nucleotide polymorphism

SOST Sclerostin gene

TCF7L1 Transcription factor 7 like 1 gene

TEAD4 TEA domain transcription factor 4

TEF Transcriptional enhancer factor

TFF Trefoil factor

TG Triglyceride

TGFB1 Transforming growth factor-β1 gene

TIMP Tissue inhibitor of metalloproteinases

TLE1 Transducin like enhancer of split 1 gene

TMEM135 Transmembrane protein 135 gene

TNC Tenascin-C gene

TNFRSF1A	Tumour necrosis factor receptor 1 gene
TNFα	Tumour necrosis factor α
TNS1	Tensin 1 gene
tRNA	Transfer ribonucleic acid
TSI	Tibial stress injury
UXT	α-class prefoldin protein
VDR	Vitamin D receptor gene
VNTR	Variable number tandem repeat
WHO	World Health Organisation
XRN2	5'-3' exoribonuclease 2

1. Chapter One - Introduction and literature review

Addendum

Contributions to Chapter 1:

Mariia Kozlovskaia:

- Literature review
- A systematic review of biomedical risk factors of Achilles tendinopathy
- Author of the chapter

Nicole Vlahovich:

- Co-author of the systematic review
- Editing of the chapter

David Hughes:

- Co-author of the systematic review
- Editing of the chapter

Kevin Ashton:

- Co-author of the systematic review
- Editing of the chapter

Publication outcome of this chapter:

Kozlovskaia, M., Vlahovich, N., Ashton, K.J. and Hughes, D.C., 2017. Biomedical Risk Factors of Achilles Tendinopathy in Physically Active People: A Systematic Review. Sports Medicine-Open, 3(1), p.20.

1.1 The role of physical activity in health

First acknowledgements of the important role of physical activity in health can be traced to ancient times and are represented by the development of medical gymnastics in China and the integration of physical activity in educational programs and establishment of the culture of athletic festivals in Greece. However, physical activity was unpopular during the Dark Ages, although it was a practical component of religion during the Middle Ages (MacAuley, 1994). Until the twentieth century, physical activity was viewed sceptically due to its potential danger to health when practised vigorously. When several epidemiological studies conducted in the early twentieth century found a positive association between physical activity and life longevity, and it regained the acknowledgement as a beneficial practice for health (Lee et al., 2012; MacAuley, 1994).

Development of research methods and accumulation of large health datasets across the world have allowed researchers to perform a thorough assessment of the current physical activity levels in different populations and their role in mortality rates from non-communicable diseases. A systematic review of long-term benefits of physical activity showed negative relationships between physical activity and weight gain or obesity, the occurrence of coronary heart disease and the risk of type 2 diabetes (Reiner, Niermann, Jekauc, & Woll, 2013). Physical activity was also negatively related to the incidence of dementia and Alzheimer's disease; however, these studies were based on weak evidence, due to the absence of longitudinal studies of people without these conditions at the baseline assessments (Reiner et al., 2013).

Regular physical activity contributes to the primary and secondary prevention of several chronic diseases, including cardiovascular disease, diabetes, hypertension and obesity (Warburton, Nicol, & Bredin, 2006). Physical inactivity causes 6-10 of the burden of disease worldwide and 9 of premature mortality (Lee et al., 2012; Physical Activity Guidelines Advisory Committee, 2008; Reiner et al., 2013). Currently, the World Health Organisation (WHO) recommends either 150 minutes of moderate-intensity aerobic physical activity or 75 minutes of vigorous-intensity physical activity throughout the week, based on strong evidence of health benefits and reductions in mortality rates (World Health Organisation, 2010).

Although the role of physical activity is proven in the maintenance of a healthy lifestyle and in the reduction in risks of various diseases, it is also accompanied by certain risks that must

be considered, in particular, musculoskeletal injuries. The incidence of these acute and overuse injuries varies depending on the type of sport or activity. This risk can be further determined by the intensity, frequency of exercise and history of previous injury (Melzer, Kayser, & Pichard, 2004). Regular physical activity at the recommended level is widely known to be beneficial to cardiovascular health (O'keefe & Lavie, 2013). However, there is a U-shaped relationship between exercise and cardiac health and morbidity. At one end of the U-curve, physical inactivity and sedentary lifestyle were associated with a higher risk of cardiovascular disease (CVD) (Myers, 2003; Richardson, Kriska, Lantz, & Hayward, 2004). In contrast, the other end of the U-curve relates to long-term excessive exercise, which may lead to cardiac damage and therefore increase risks of sudden cardiac arrest and cardiovascular disease mortality (O'keefe & Lavie, 2013). The majority of the sudden deaths associated with training can be explained by an existing cardiovascular abnormality (Melzer et al., 2004). Overall, health risks related to physical activity are mainly associated with the level of its intensity, not the physical activity itself.

1.2 Physical activity in Australia

According to the 'Australian Burden of Disease Study', physical inactivity is one of the top five factors contributing to the burden of disease (Australian Institute of Health and Welfare, 2016b). Physical activity guidelines developed by the Australian Department of Health comprise recommendations of the WHO and are distinguished for four age groups (Department of Health, 2012). Whilst the first three age groups describe the young population, under 18 years, the fourth age group is comprised of individuals aged between 18 and 64 years. For this (fourth) age group, it is recommended to accumulate 150 to 300 minutes of moderate intensity physical activity or 75 to 150 minutes of vigorous intensity physical activity per week.

According to a report from the National Health Survey (2014-2015), 57.7 of Australian men and 53.3 of Australian women aged between 18 and 64 years met recommended physical activity guidelines (Australian Bureau of Statistics, 2016a). Approximately 79 of Australian adults participated in physical activity at least once per week with similar levels of participation for men and women across the life stages (Australian Sports Commission, 2016). According to the most recent report from the Australian Bureau of Statistics, the three most popular physical activities among Australians were walking for exercise (19.2), fitness or gym (17.4) and running or jogging (7.4) (Australian Bureau of Statistics, 2015).

1.3 Health benefits and risks of running and jogging

Running or jogging is a low cost and an easily implemented activity with proven health benefits involving improvements in aerobic fitness and cardiovascular function (Oja et al., 2015). A longitudinal study, the 'Copenhagen City Heart Study', showed that jogging was associated with a significantly lower all-cause mortality over a 35-year follow up period (Schnohr, Marott, Lange, & Jensen, 2013). Another large longitudinal study of 55,137 adults showed that people who incorporated running in their training had 30 and 45 lower risks of all-cause and CVD mortality, respectively (Lee et al., 2014). Interestingly, running at lower doses per week, and slower speeds still was associated with reduced mortality risks, and persistent running over time was associated with mortality reduction (Lee et al., 2014).

Running is a sport with a relatively high incidence rate of lower extremity injuries, varying between 19 and 79 (van Gent et al., 2007). Approximately 80 of running-related injuries are overuse injuries mainly concerning lower limb, particularly the Achilles tendon, knee and foot (Walther, Reuter, Leonhard, & Engelhardt, 2005). Acute injuries that occurred while running are less common. A systematic review of the main running-related injuries reported that the most common injuries were: medial tibial stress syndrome (known as a bone stress injury) (13.6-20), Achilles tendinopathy (9.1-10.9) and plantar fasciitis (4.5-10) (Dias Lopes, Hespanhol Junior, Yeung, & Pena Costa, 2012). Overuse injuries have multifactorial aetiology; the development of these injuries depends on intrinsic (non-modifiable) factors, including sex, age, systemic disorders and genetic predisposition, and also extrinsic (modifiable) factors such as training loads and lifestyle-related factors. Risk factors of overuse injuries have been classified and reported in various prospective studies and systematic reviews, however, the interrelationship between extrinsic and intrinsic risk factors and their relative contribution to the development of overuse injuries remains unclear.

1.3.1 Risk factors of running-related injuries

1.3.1.1 Physical characteristics of runners

Age is a non-modifiable factor and has been studied in relation to the injury risk in multiple cohorts. Two systematic reviews indicated age as a risk factor of running injuries with the conflicting evidence; some studies reported higher age as a risk factor and other studies referred to the greater age as a protective factor against injuries (van der Worp et al., 2015;

van Gent et al., 2007). Sex is an intrinsic factor which also demonstrated conflicting associations with development of certain types of overuse injuries, but no significant association with injury occurrence in total (Hart, 1994; van der Worp et al., 2015; van Gent et al., 2007; Walter, Hart, McIntosh, & Sutton, 1989). Evaluations of body mass index as a risk factor of running injuries also showed inconsistent results, these results possibly were affected by the relatively small numbers of overweight and obese runners in the studies (Van Mechelen, 1992). However, a study of risk factors of running-related injuries in novice runners showed that the increase of the body mass index by 1 kg/m^2 was associated with an increased hazard risk of the running-related injury in males (Buist, Bredeweg, Lemmink, Van Mechelen, & Diercks, 2010).

1.3.1.2 Anatomic and biomechanical characteristics of runners

Certain alignment defects may be important factors in developing overuse injuries (Van Mechelen, 1992). Varus or valgus knee alignment, significant leg length discrepancy, high Q-angle and foot anomalies were associated with higher risk of injuries in several studies (Korpelainen, Orava, Karpakka, Siira, & Hulkko, 2001a; Taunton et al., 2003; Wen, Puffer, & Schmalzried, 1998; Williams Iii, McClay, & Hamill, 2001). Muscle weakness or imbalance in the strength of muscle groups may also increase the risk of injuries in runners (Fredericson et al., 2000; Niemuth, Johnson, Myers, & Thieman, 2005). Running biomechanics are dictated by lower limb anatomy, particularly joints of the foot and ankle (Dugan & Bhat, 2005). It has been shown that altered biomechanics played a major role in the development of the exercise-related lower leg pain (Willems, Witvrouw, De Cock, & De Clercq, 2007). A study that analysed biomechanical and anthropometric variables of groups of injured and uninjured runners found that runners whose stride patterns incorporated lower levels of impact forces had reduced risk of injuries (Hreljac, Marshall, & Hume, 2000). Several studies showed that retraining of the running gait and cadence led to a significant reduction in pain (Barton et al., 2016; Esculier et al., 2017; Roper et al., 2016).

1.3.1.3 Training characteristics and running habits

A review of running-related musculoskeletal injuries showed that overtraining and the presence of the previous injury were the most significant predictors of the overuse injuries (Hart, 1994). Two large studies on recreational runners reported that running over 64 km per week was a risk factor for running-related injuries (Macera et al., 1989; Walter et al.,

1989). Weekly frequencies of running sessions were also studied as a risk factor and demonstrated contradictory results: one study reported that frequencies of three to seven times per week for men and seven times per week for women were associated with risk of injury (Walter et al., 1989). However, another study demonstrated that one session per week was a risk factor for injury for women (Taunton et al., 2003). A systematic review of determinants of lower extremity running injuries in distance runners reported conflicting evidence for the association of training increase with the risk of injury (van Gent et al., 2007). A previous injury occurring in the preceding 12 months has been identified as the main predictor of running injury. This association might be due to incomplete recovery from the previous injury (Saragiotto et al., 2014; Walter et al., 1989). Stretching is another factor that has been studied in association with running-related injuries and has shown ambiguous results (Hreljac, 2005). A critical review investigated associations between stretching and injury risk and reported that prospective studies that showed beneficial effects of stretching incorporated other interventions, whereas prospective studies that investigated stretching alone did not find any significant difference in injury rates. Moreover, the trend was towards higher injury rates in people who stretch before exercise (Shrier, 1999).

Running shoes provide support and cushioning and play an essential role in shock absorption while running and hence, injury prevention (Van Mechelen, 1992). In 2008, a systematic review of commonly prescribed running shoes with heel elevation demonstrated that these prescriptions were not evidence-based (Richards, Magin, & Callister, 2009). Advancements in shoe technology and running biomechanics research led to the expansion of barefoot running and development of minimalist footwear (Rixe, Gallo, & Silvis, 2012). Thus, it was hypothesised that minimalist shoes would allow runners to disperse impact forces more efficiently than runners in conventional shoes. A clinical trial of full and partial minimalist footwear on 103 recreational runners showed that injury rates among runners wearing these minimalist types of footwear were higher than among runners wearing neutral footwear (Ryan, Elashi, Newsham-West, & Taunton, 2014). However, another study of 1,332 American male soldiers compared injury risks between soldiers wearing traditional and minimalist running shoes and did not identify any significant difference (Grier et al., 2016). As the midsole thickness and other shoe characteristics affect running biomechanics and therefore, injury risk (Law et al., 2018), the choice of the running footwear may be an important contributing factor. Orthotics influence the biomechanical control of the foot, aiming to improve running and provide additional cushioning for the foot (McKenzie,

Clement, & Taunton, 1985; Nigg, Nurse, & Stefanyshyn, 1999). Appropriately prescribed orthotics may assist in injury prevention and are often considered to have a role in the treatment of the injury. However, poorly constructed orthotics have been linked to increased incidence of injury (McKenzie et al., 1985).

1.3.1.4 Lifestyle habits

Certain lifestyle habits, including smoking status and alcohol consumption, may have a negative impact on physical activity and running. A study of smoking habits among runners showed that 50 of smokers that participated in the study had given up smoking since starting to run (Ortega & Aguilar-Blanco, 2006). Although smoking is not a common habit among physically active people, the majority of the studies that investigated risk factors of running-related injuries found a negative association between smoking and injuries (Kraemer et al., 2012; Van Middelkoop, Kolkman, Van Ochten, Bierma- einstra, & Koes, 2008). According to a systematic review on risk factors of running-related injuries, there was limited evidence of the association between alcohol consumption associated and injuries (van Gent et al., 2007).

Studies have shown the most prevalent running-related injuries are medial tibial stress syndrome, Achilles tendinopathy and plantar fasciitis (Dias Lopes et al., 2012). Due to many contradictory conclusions about risk factors and their impact on running-related injuries in the literature, the two most prevalent injuries of lower limbs have been selected for a detailed review of their risk factors: Achilles tendinopathy and bone stress injuries.

1.4 Achilles tendinopathy

1.4.1 Achilles tendon anatomy

Anatomically, tendons are the structures which connect muscles and bones and transmit the force from muscle to bone and enable joints to move. Tendons are essential to accommodate sustained loads (Sharma & Maffulli, 2005). The Achilles tendon is the largest and the strongest tendon in the human body and plays a fundamental role in locomotion. It is formed by the confluence of gastrocnemius and soleus muscles and connects to the calcaneus (Figure 1.1). Tenocytes and tenoblasts constitute 95 of cells within the tendon. Tenocytes are elongated fibroblast type cells and responsible for collagen secretion. Tenoblasts are variable in shapes and arranged in long parallel chains (Longo, Ronga, & Maffulli, 2009b). Collagen constitutes 90% of the tendon's extracellular matrix. The most common is the collagen type I and the second common is collagen type II. Overall, collagen accounts for 65-80 of the dry tendon mass, with 1-2 of the dry tendon mass being elastin (Padhiar et al., 2010). The principal role of collagen fibres is to resist tension and at the same time, allow for a certain degree of compliance. A hierarchical structure of the tendon resolves the contradiction between these tendon's functions (Figure 1.2) (Benjamin, Kaiser, & Milz, 2008). Tendons have significantly less vascular supply comparing to muscles and therefore have 7.5 times lower oxygen consumption than muscles (Benjamin et al., 2008; Sharma & Maffulli, 2005). Laser Doppler flowmetry showed that blood flow is distributed evenly along the tendon and is lower near the insertion to calcaneal (Kader, Saxena, Movin, & Maffulli, 2002). Innervation of the tendon is superficial, and nerves follow blood vessels along the tendon (Padhiar et al., 2010).

Figure 1.1 Achilles tendon anatomy (adapted from (Barfod, 2014)).

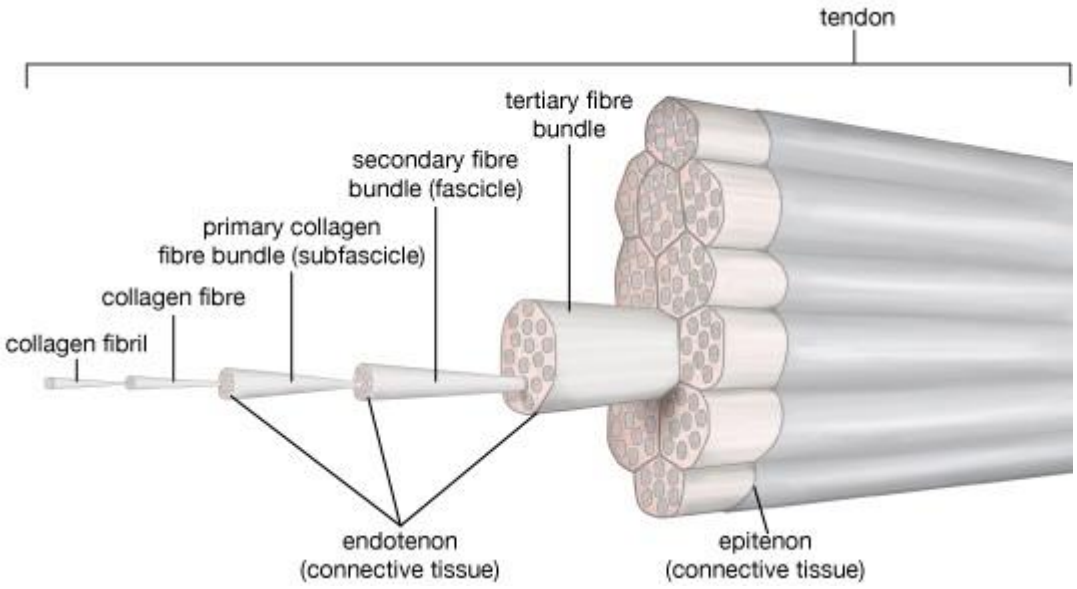


Figure 1.2 Hierarchical organisation of tendon structure (adapted from (Encyclopaedia Britannica, 2017b)).

1.4.2 Achilles tendon pathology

Achilles tendinopathy is the most common clinical diagnosis of Achilles disorders and accounts for 55-65 of Achilles tendon injuries, with the highest incidence among runners, track and field athletes, volleyball, tennis and soccer players (J rvinen, Kannus, Maffulli, & Khan, 2005). According to a systematic review on running-related musculoskeletal injuries in 3,276 runners, the prevalence of Achilles tendinopathy has been estimated between 6.2 and 9.5 (Dias Lopes et al., 2012). However, other studies of runners have demonstrated a prevalence rate of Achilles tendinopathy varying between 11 and 24 (Wasielewski & Kotsko, 2007). The term 'tendinopathy' is an umbrella term for the description of tendon conditions encompassing pain, swelling and impaired performance (Maffulli, 1998; Padhiar et al., 2010). Whilst inflammatory cytokines have been observed in the pathological tendons, their expression seemed to differ from a traditional inflammatory response. Some studies demonstrated that the expression of these cytokines may be a response of tendon cells to mechanical stimuli (Kjaer, Bayer, Eliasson, & Heinemeier, 2013; Li et al., 2004). This also may be a sign of imbalance between synthesis and degradation processes, which may lead to tendon disorganisation. Clinically, Achilles tendinopathy develops as a result of strenuous physical activity with a continuous overload of the tendon and is typically characterised by pain and swelling around the tendon and morning stiffness. (Padhiar et al., 2010). As load that is applied to the tendon can be both anabolic and catabolic, repetitive energy storage and release, and excessive compression may be key factors in the development of tendinopathy. Even though overload is the main pathological component, it is modulated by other factors including genetics, sex, age and biomechanics (Cook & Purdam, 2009).

A continuum model of tendon pathology, which was proposed in 2009, has three stages: reactive tendinopathy, tendon disrepair and degenerative tendinopathy (Figure 1.3) (Cook & Purdam, 2009). The first stage develops as a reaction to acute overload and can be distinguished by changes in and extracellular matrix structure and cell shape, yet collagen integrity is maintained. The second stage is described as an attempt at tendon healing through the increased production of collagen and proteoglycans and is characterised by greater disorganisation of the matrix. The tendon itself may be swollen, and its vascularity increased. The third stage is characterised by potentially irreversible changes in cell and matrix condition, such as tenocyte apoptosis and matrix disorder (Cook & Purdam, 2009).

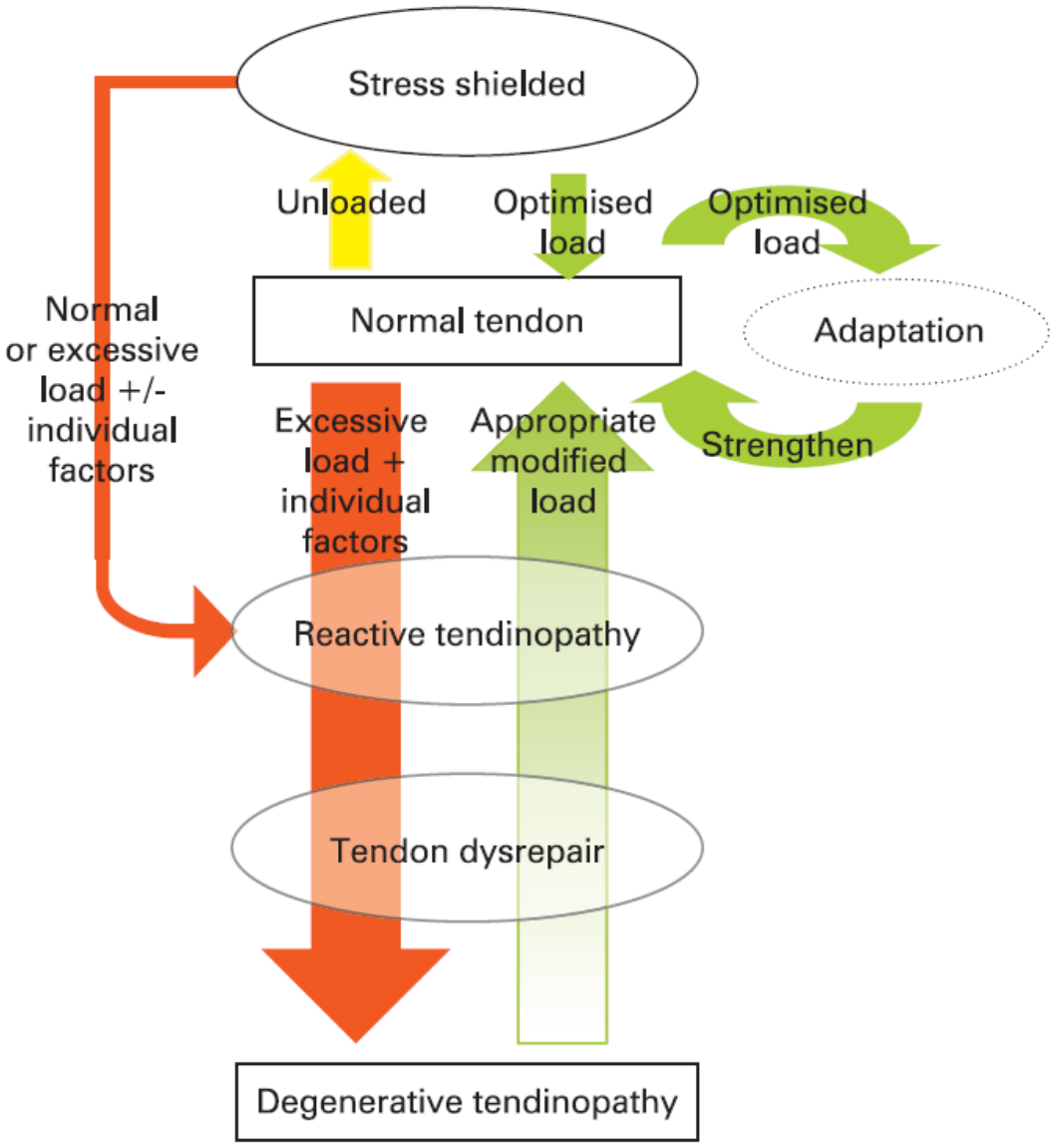


Figure 1.3 Continuum model of tendon pathology.

Although Achilles tendon rupture can be a result of untreated Achilles tendinopathy, typically spontaneous Achilles tendon ruptures occur without any preceding symptoms (Cook & Purdam, 2009). Achilles tendon ruptures may be associated with participation in sports. Histopathologic analysis of ruptured tendon showed obvious degenerative changes which might be caused by a sedentary lifestyle. Therefore, spontaneous tendon ruptures can be caused by asymptomatic degeneration of the tissue and following sudden movement (Figure 1.4) (J rvinen et al., 2005).

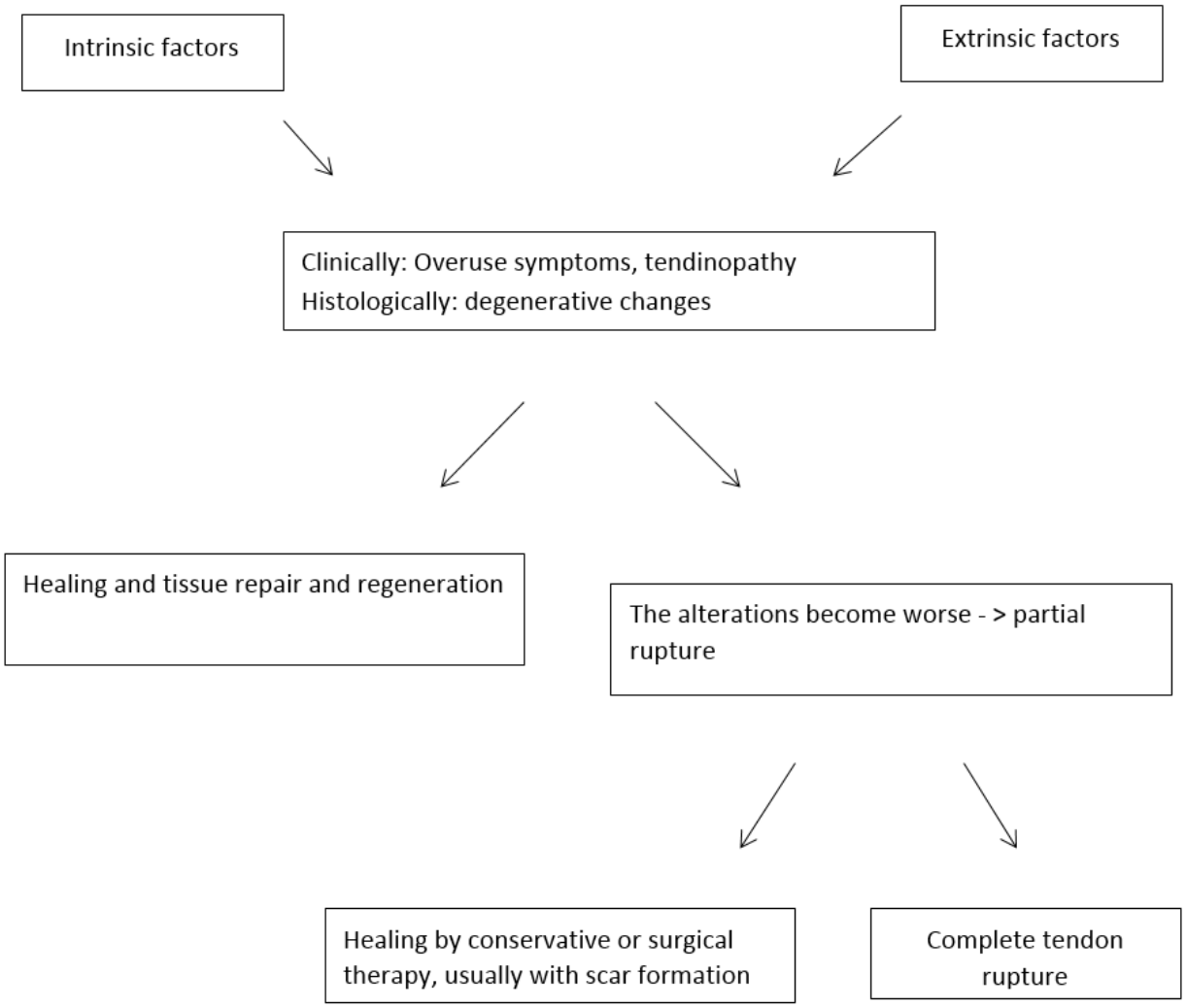


Figure 1.4 Pathophysiologic mechanisms of an Achilles tendon injury development (adapted from (Järvinen et al., 2005)).

For treatment decisions, tendon pathology stages may be divided into two groups: reactive/early tendon disrepair and late tendon disrepair/degenerative. Conservative treatment and modification of the load may be suggested for the first stage, whereas the second stage may require surgery after failed conservative interventions (Cook & Purdam, 2009). However, the recovery process in the tendon is slow due to poor blood flow and slow synthesis of collagen (Kader et al., 2002). Overall, tendon's response to treatment is very slow, both in regards to improving load capacity and resolving pain (Cook & Purdam, 2009). The key priority in the management of Achilles tendinopathy is therefore, prevention.

Aetiological factors of Achilles tendinopathy are also divided into intrinsic and extrinsic, with the interaction leading to the symptoms and further progress of the tendinopathy. Intrinsic

risk factors include demographic non-modifiable factors (sex, age and height), genetic susceptibility; and local anatomic factors: leg length discrepancy, malalignment and decreased flexibility. Extrinsic factors comprise weight and body mass index (BMI), therapeutic agents (corticosteroids, antibiotics), poor environmental conditions (cold weather, icy surface), and sport-related factors, involving training patterns, technique and equipment (J rvinen et al., 2005; Padhiar et al., 2010).

1.4.3 Modifiable risk factors of Achilles tendinopathy

1.4.3.1 Weight characteristics, body mass index and lipid profile.

A case-control study of 60 patients with mid-portion Achilles tendinopathy and 60 uninjured controls matched by age, sex and BMI showed that patients with Achilles tendinopathy had higher levels of triglycerides (TG), lower levels of high-density lipoprotein cholesterol (HDL-C) and higher TG:HDL-C ratio. These lipid profiles are typical for insulin resistance syndrome and usually described as dyslipidemia. This finding suggests that serum lipids may be involved in the development of Achilles tendinopathy (Gaida, Alfredson, Kiss, Bass, & Cook, 2010; Gaida et al., 2009). Another study investigated fat distribution on male and female patients with asymptomatic Achilles tendinopathy. Ultrasound examinations of Achilles tendon in 298 participants identified a higher pathology rate among men than women (17/127, 13 versus 8/171, 5). Males with asymptomatic Achilles tendon pathology had elevated waist to hip ratio, were older, had higher central/peripheral fat mass and larger waist circumferences (above 83cm) compared to individuals without tendon pathology. This pattern of fat tissue distribution around the abdominal area is usually associated with metabolic syndrome and may be linked to the previously reported association between dyslipidemia and Achilles tendinopathy. Surprisingly, women with asymptomatic Achilles tendon pathology had lower central to peripheral fat mass ratios compared to women without pathology, which could be explained by the effect of oestrogen on body fat distribution, a factor that was not investigated in this study (Gaida et al., 2010). A recent study comparing patients with Achilles tendinopathy aged over 65 years old to matching uninjured controls showed a significantly higher prevalence of diabetes, higher BMI and a higher level of sport participation in the injured group (Abate, Salini, & Schiavone, 2015). These findings are consistent with other studies that indicate susceptibility in those with chronic comorbidities. The types of exercises undertaken by the participants (speed

walking, jogging and tennis) were relatively high-risk in nature and may have been beyond the loading capacity of the tendon in ageing individuals (Abate et al., 2015). The link between an adverse lipid profile and tendinopathy could be explained by the adipokine modulation of certain enzymes' production, which are important for tenocyte functioning. Additionally, chronic low-grade inflammation, which is typical for obesity, may affect the tendon healing process. At the same time, tendon healing is also disrupted by low concentrations of immune cells that tend to migrate into adipose tissue (Abate, Oliva, Schiavone, & Salini, 2012). A systematic review of 17 articles on the link between lipid profile and tendon health showed a strong association between tendon pathology and high lipid parameters (Tilley, Cook, Docking, & Gaida, 2015).

1.4.3.2 Running habits and training characteristics

A review of running-related risk factors of Achilles tendon injuries showed that running distances and number of years trained had contradictory results in different studies but overall did not demonstrate any significant effects on the development of Achilles tendon injuries (Lorimer & Hume, 2014). A systematic review of running-related musculoskeletal injuries reported that excessive load generated in the gastrocnemius and soleus muscles was the main stimulus for the development of Achilles tendinopathy (Dias Lopes et al., 2012).

The majority of runners prefer to run on hard terrains, such as bitumen and cement. Soft terrains, particularly sand, were reported as a risk factor of Achilles tendinopathy development, whereas asphalt had a protective effect against this injury; this effect might be related to the higher range of motion of the foot during running on the sand (Knobloch, Yoon, & Vogt, 2008).

Correction with orthotics alters the biomechanics of the foot and helps to relieve heel pain, reduce strain on the Achilles tendon and reduce the risk of the Achilles tendon injury (Farris, Buckeridge, Trewartha, & McGuigan, 2012; Kader et al., 2002). Shoe adaptations with special insoles may help to prevent Achilles tendon injuries (Peters, werver, Diercks, Elferink-Gemser, & van den Akker-Scheek, 2015). Stretching is typically thought to be a preventive method against sports injuries; however, there is no conclusive evidence for its positive or negative effect (Peters et al., 2015). McCrory et al. showed that runners with Achilles tendon injuries were less likely to incorporate stretching in their training routine, but

unfortunately the study survey didn't include questions about time relationship between stretching practice and injury occurrence (McCrory et al., 1999).

1.4.3.3 Lifestyle habits

Smoking is a well-known risk factor for a wide range of diseases, such as multiple types of cancer and CVD (Erhardt, 2009; World Health Organization, 2006), and was considered as a potential risk factor of Achilles tendinopathy in a number of studies (Kraemer et al., 2012; Owens et al., 2013). However, in these studies results were contradictory, as Kraemer et al. found a negative association between Achilles tendinopathy and positive smoking status, suggesting that physically active people were less likely to have a smoking habit. Owens et al. reported no significant association between smoking status and Achilles tendinopathy. Although studying smoking as a risk factor for exercise-related disorders or injuries is complicated (due to negative association between smoking and physical activity), this risk factor should be taken into account in relation with tendon injuries.

A study of 450 cases of Achilles tendinopathy in a large military cohort found that moderate alcohol use was associated with an increased risk of Achilles tendinopathy. However, this finding was weak, as the magnitude of the OR was modest (OR 1.33, CI: 1.00-1.74) and this association could be explained by the alcohol-related risk-taking behaviour which may lead to an injury, and alcohol's influence on metabolic and inflammatory factors that may contribute to the tendon pathology (Owens et al., 2013).

1.4.3.4 Medications affecting tendon tissue

Corticosteroid injections were introduced as a therapy for inflammatory conditions in 1950. At that time Achilles tendinopathy was thought to be accompanied by inflammation and therefore treated by corticosteroid injections. However, further research revealed adverse effects of local steroid injections. The first comprehensive literature review of direct corticosteroid injections into Achilles tendon and also peritendinous injections revealed that intratendinous injections should be abandoned due to the high incidence of subsequent tendon rupture (Mahler & Fritschy, 1992). The main conclusion of the review was that local corticosteroid injections can mask symptoms of the degenerative state of the tendon and therefore expose the tendon to the trauma (Mahler & Fritschy, 1992). Although chronic

Achilles tendinopathy is a non-inflammatory disorder, corticosteroid and nonsteroidal anti-inflammatory drug injections are proposed as part of the conservative management of Achilles tendinopathy (Longo et al., 2009b). The controversy regarding the efficacy of these injections and possible adverse effects raises questions over the use of corticosteroids and makes it a considerable risk factor of Achilles tendon injuries.

Anabolic steroids are used to promote muscle growth and tissue repair, but adverse side effects were demonstrated by several studies on animals (Inhofe, Grana, Egle, Min, & Tomasek, 1995). A review by Laseter et al. showed that anabolic steroids intake had significant side effects, including dysplasia of tendon fibrils (Laseter & Russell, 1991).

Fluoroquinolone antibiotics were introduced in the 1980s and are used to treat respiratory infections caused by Gram-negative and anaerobic bacteria. Fluoroquinolone antibiotics inhibit fibroblast metabolism through the stimulation of their matrix-protease activity and also reduce collagen and proteoglycan synthesis (Williams, Attia, Wickiewicz, & Hannafin, 2000). In Denmark, a population-based cohort study of fluoroquinolone antibiotics administration demonstrated that individuals had three times higher risk of Achilles tendon rupture within 90 days of using fluoroquinolone antibiotics when compared to the background population (Sode, Obel, Hallas, & Lassen, 2007). An Italian population case-control study used health services databases and investigated the influence of fluoroquinolone antibiotics and corticosteroids on the development of different Achilles tendon injuries. This study showed that patients treated with fluoroquinolone antibiotics are at 1.7-fold increased risk of Achilles tendon injuries as a whole and in particular at a 4.1-fold increased risk of Achilles tendon ruptures. Moreover, simultaneous exposure to corticosteroids may multiply the risk of Achilles tendon injuries by 10 times (Corrao et al., 2006). A literature review of Achilles tendon injury cases among patients prescribed fluoroquinolone antibiotics showed strong evidence of the adverse effect on the tendon (Khaliq & hanel, 2003).

1.4.4 Non-modifiable risk factors of Achilles tendinopathy

1.4.4.1 Physical characteristics of recreational runners with Achilles tendinopathy

Physical characteristics, such as age and sex, are considered as non-modifiable traits, and commonly reported and investigated as factors of multifactorial conditions, such as Achilles tendinopathy. Thus, ageing increases collagen and decreases glycosaminoglycan

concentrations in tendons (Longo et al., 2009b), therefore reducing tendon elasticity and forcing an increased load on muscles. However, a cross-sectional study of the risk factors of Achilles tendinopathy in 178 master athletes (aged over 35) carried out by Longo et al. did not find any association between age, sex, weight or height and the development of Achilles tendinopathy (Longo et al., 2009a). A systematic review of the pathogenesis of Achilles tendinopathy analysed data from 68 articles. In 19 of these articles, age was reported as a factor affecting the tendon matrix in different ways (Magnan, Bondi, Pierantoni, & Samaila, 2014). In a retrospective study of 2002 running injuries, males younger than 34 years old were less likely to develop Achilles tendinopathy (Taunton et al., 2002). Additionally, Hirshmuller et al. showed that older age was a risk factor for midportion Achilles tendinopathy among 634 long-distance runners (Hirschm ller et al., 2012).

Men report Achilles tendinopathy more often than women. However, considering differences in physical activities between the sexes, it is hard to evaluate sex as an independent factor (Magnan et al., 2014). A study of high school runners aged 13-18 failed to find any association between biological sex and the development of Achilles tendinopathy (Tenforde et al., 2011).

1.4.4.2 Biomechanical and anatomic characteristics

Biomechanical characteristics of muscles and joints, such as strength, flexibility and range of motion, are frequently discussed as being intrinsic risk factors for Achilles tendinopathy. A systematic review of biomechanical risk factors for Achilles tendinopathy in runners identified a high foot arch as a protective factor against Achilles tendinopathy, whereas large peak braking force was the variable with the most negative effect on the tendon health (Lorimer & Hume, 2014). A study of 89 runners showed that weak plantar flexion and lower plantar flexion average power were typical characteristics of runners with Achilles tendinopathy (McCrory et al., 1999). Several studies demonstrated that malalignment and hyperpronation may contribute to the development of Achilles tendinopathy (J rvinen et al., 2005).

1.4.4.3 Genetic polymorphisms associated with Achilles tendinopathy

Genetic variations are sequence alterations in deoxyribonucleic acid (DNA) among individuals, these may account for differences in phenotype and also health/disease status. One of the most commonly investigated types of genetic variation is the single nucleotide polymorphism (SNP), a variation in a single nucleotide located at a specific position in the DNA (Wright, 2005). Genetic variations occurring in more than 1 of a population are considered useful polymorphisms for genetic linkage analysis. To date, the majority of genetic studies have employed a candidate gene approach to find genetic polymorphisms in genes that might influence tendon structure and its development and therefore either predispose to or protect from, the development of Achilles tendinopathy. Genetic predisposition has long been proposed as a contributor to the development of Achilles tendinopathy. Early studies assessed the genetic association between ABO blood type and tendon injuries, described in several investigations of Finnish and Hungarian patients (Leppilahti, Puranen, & Orava, 1995; Maffulli, Reaper, Waterston, & Ahya, 2000). However, further research did not support this link with ABO blood group (Mokone, Schwellnus, Noakes, & Collins, 2006). However, other genes located in the same region of chromosome 9q as the ABO blood group gene were implicated, as further detailed below. Furthermore, Table 1.1 summarizes all genes and genetic polymorphisms that have currently been investigated in association with Achilles Tendinopathy.

Examination of chromosome region 9q revealed the presence of two tendon-related genes: tenascin-C (*TNC*) and collagen, type V, alpha 1 (*COL5A1*) (Mokone et al., 2005; Mokone et al., 2006). Studies found an association of polymorphic variants of *TNC* and *COL5A1* genes with the development of Achilles tendinopathy. Tenascin-C regulates cell-matrix interaction in the tendon. *TNC* contained a polymorphism in intron 17 where the number of GT dinucleotide repeats ranged from 3 to 21, and 95 of the alleles contained 12 to 17 GT repeats. In this range, the most common alleles contained 12 and 14 repeats, and the least common alleles were of 13 and 17 repeats. Those runners, who were homozygous or heterozygous for the underrepresented alleles of *TNC,* were 6.2 times less likely to develop an Achilles tendon injury (Mokone et al., 2005) indicating that these variants of *TNC* may be protective from Achilles tendinopathy. *COL5A1* encodes the pro-a1 (V) chain of the type V collagen. Two variants of the *COL5A1* gene were associated with an increased risk of Achilles tendinopathy. In contrast, the third variant was underrepresented in runners with Achilles tendinopathy and therefore, was associated with a reduced likelihood of Achilles

tendinopathy development (Mokone et al., 2006). A follow-up study published by the same research group from South Africa in collaboration with researchers from Australia replicated previously shown associations linking *COL5A1* gene variants to Achilles tendinopathy in both Australian and South African populations (September, 2009). These studies demonstrate the relevance of the *TNC* and *COL5A1* gene polymorphisms in genetic predisposition to Achilles tendinopathy.

Other studies have assessed the contribution of haplotypes constructed from candidate polymorphisms in an effort to find interactions between genes or gene products. A collaborative study investigated collagen, type XXVII, alpha 1 (*COL27A1*) gene polymorphisms in conjunction with *TNC*, as *COL27A1* is located in the same region of chromosome 9q as the *TNC* gene. Type XXVII collagen, encoded by *COL27A1*, is responsible for the critical structural framework and tensile strength of the interstitial matrices. Although there were no significant associations between the polymorphisms in *COL27A1* and Achilles tendinopathy, a GCA haplotype (a set of genetic variants located on a single chromosome), constructed from one *COL27A1* polymorphism (rs946053) and two *TNC* polymorphisms (rs13321, rs2104772), showed a significant association with Achilles tendinopathy (Saunders et al., 2013). The study of *COL5A1* and *MIR608* polymorphisms in South African and Australian cohorts, found that a variant of *COL5A1* contains a putative polymorphic micro-RNA binding site. The results showed that polymorphisms rs71746744, rs16399, rs1134170 in *COL5A1*, and a polymorphism rs4919510 in *MIR608*, which encodes a small microRNA 608, were all independently associated with Achilles tendinopathy, suggesting a role for these four variants in the messenger ribonucleic acid (mRNA) stability and the consequent type V collagen synthesis (Abrahams, Laguette, Prince, & Collins, 2013). A follow-up study of the same polymorphisms in a British cohort did not find any independent association between studied *COL5A1, MIR608* or *IL-1B* (interleukin 1β) variants and Achilles tendinopathy as was shown in Australian and South African cohorts (Brown et al., 2016). However, an inferred allele combination constructed from *COL5A1* SNPs rs12722, rs3196378 and rs71746744 was associated with the risk of Achilles tendon pathology (Brown et al., 2016).

A further study conducted on South African study participants analysed polymorphisms in the gene encoding for matrix metalloproteinase 3 (*MMP3*), which is involved in the regulation of extracellular matrix homeostasis. Three of the investigated polymorphisms in the *MMP3* gene (rs679620, rs591058, rs650108) showed strong association with Achilles tendinopathy

and the most underrepresented haplotype in patients with Achilles tendinopathy indicated that this variant was protective against Achilles tendinopathy (Raleigh, 2009). Moreover, since *MMP3* genotyping had been done on the same cohort as *COL5A1* reported by September et al. (September, 2009), this study presented allelic combinations of *MMP3* rs679620 and *COL5A1* rs12722, which are associated with a lower risk of Achilles tendinopathy. Type V collagen is a substrate for MMP3, hence genetic variants in *COL5A1* and *MMP3* genes could account for differences in the interactions between the proteins (Raleigh, 2009). Further investigation of *MMP3* gene's polymorphisms showed that an inferred haplotype of four SNPs (rs3025058, rs679620, rs591058 and rs650108) was associated with Achilles tendinopathy in an Australian cohort (Gibbon et al., 2016).

Polymorphisms in *COL12A1* (rs240736, rs970547) and *COL14A1* (rs4870723, rs1563392) were investigated as they both encode for proteins involved in the biological processes of fibrillogenesis and, like tenascin-C, in the modulation of the tendon response to mechanical stress (September et al., 2008). Additionally, one study of two cohorts from South Africa and Australia investigated polymorphisms in the three genes coding for type XI collagen that is homologous to type V collagen in function and structure (Hay et al., 2013). Type XI collagen is usually expressed in cartilage, but also in developing tendons. Several polymorphisms (*COL11A1* rs3753841 and rs1676486, *COL11A2* rs1799907) have been associated with lumbar disc herniation and rheumatoid arthritis. Whilst none of the polymorphisms were independently associated with Achilles tendinopathy, the construction of a pseudohaplotype consisting of three polymorphisms in *COL11A1* and the *COL5A1* polymorphism rs71746744, revealed a significant association with Achilles tendinopathy. It was hypothesized that the interaction of genes encoding for type V and XI collagens could modulate the risk of Achilles tendinopathy, allowing for the possibility that the effects of type XI collagen variants in the developing tendon may affect the structural or functional properties of the mature tendon (Hay et al., 2013).

Genes involved in tendon homeostasis and particularly in the extracellular matrix have also been investigated for their association with Achilles tendinopathy risk (El Khoury et al., 2013). The ADAMTS (A Disintegrin And Metalloproteinase with ThromboSpondin motifs) is a family of proteinases, which are involved in the extracellular matrix homeostasis, and reported to be more highly expressed in pathological tendons than in healthy tendons. Tissue inhibitor of metalloproteinases (TIMP) inhibits the actions of MMPs and ADAMTS. Previously studied cohorts were genotyped for the *ADAMTS2* rs1054480, *ADAMTS5*

rs226794, *ADAMTS14* rs4747096, *ADAM12* rs3740199 and *TIMP2* rs4789932 gene variants. Researchers found a significant association between the rs4789932 *TIMP2* variant and Achilles tendinopathy. The balance between TIMPs and MMPs could be a contributing factor for Achilles tendinopathy development (El Khoury et al., 2013). An attempted replication of the study of *MMP3* and *TIMP2* gene variants in a British cohort showed that a gene variant in *TIMP2* rs4789932 was associated with a reduced risk of Achilles tendon pathology in males (El Khoury, Ribbans, & Raleigh, 2016). The continued investigation of proteins involved in the tendon structure included research of fibrillin and elastin for their role in elasticity, strength and flexibility of tendons. The polymorphisms *FBN2* rs331079 and *ELN* rs2071307 were studied in Australian and South African cohorts, and the GG genotype in rs331079 was overrepresented in the group with Achilles tendinopathy, indicating an association between fibrillin and the injury (El Khoury et al., 2015). Another study conducted on the same cohorts investigated the contribution of genes encoding growth factors that play an important role in tendon growth and homeostasis. *TGFB1* and *GDF5* (encoding for transforming growth factor-β1 and growth/differentiation factor-5, respectively) were selected as candidates as these proteins had been shown to increase mechanical strength after gene transfection in Achilles tendon in experimentally injured animals (Rickert et al., 2005). This study showed a significant association of Achilles tendinopathy with *GDF5* rs143383. However, no association with *TGFB1* rs1800469 was identified (Posthumus et al., 2010).

A thorough analysis of polymorphisms in *COL5A1* identified this gene as one of the most likely predisposing factors for Achilles tendinopathy. However, several studies, investigating polymorphisms in genes encoding for proteins interacting with type V collagen, showed that it is important consider possible connections and pathways, whose interactions may be disrupted and therefore alter collagen structure and its functionality and lead to the increased or decreased risk of Achilles tendinopathy.

Candidate genes involved in processes such as tendon turnover and inflammation have also been considered as possible genetic risk factors for Achilles tendinopathy. The SNPs in cytokine genes, which have been shown to be upregulated in tendinopathy and mechanically loaded tendon, were also investigated. Interleukin-1β (IL-1β) induces inflammatory mediators that upregulate the expression of proteins involved in the degradation of the tendon extracellular matrix such as MMPs which target type V collagen. The IL-1β receptor antagonist (IL-1ra) is encoded by the *IL1RN* gene and its variable number

tandem repeat (VNTR) rs2234663 polymorphism has been previously associated with gastrointestinal diseases (Mansfield et al., 1994), osteoporotic fractures (Langdahl, L kke, Carstens, Stenkj r, & Eriksen, 2000) and atherosclerosis (Olofsson et al., 2009). The genetic variants in *IL1B* (rs1143627 and rs16944) have been implicated to the increased expression of the *IL1B* gene (Landvik et al., 2009), and interleukin-6 (IL-6) was found to be linked to the tenocyte apoptosis, which is typical for tendinopathy. Thus, IL-1β and IL-6 may also affect *COL5A1* gene expression (September et al., 2011). A SNP in *IL6* (rs1800795) was previously shown to alter the *IL6* expression (Fishman et al., 1998), which may lead to the increased tenocyte apoptosis and therefore, potentially increase the risk of the development of Achilles tendinopathy. In total, the study investigated four polymorphisms in *IL1B*, *IL1RN* and *IL6*, although none of these polymorphisms were associated with the Achilles tendinopathy diagnosis in either of the South African and Australian population groups studied. However, inferred allele combinations constructed from previously studied *COL5A1* polymorphisms and *IL1B, IL6* and *IL1RN* VNTR polymorphisms were associated with an increased risk of Achilles tendinopathy in combined groups (September et al., 2011). This study concluded that genetic polymorphisms contributing to the changes in inflammatory pathways may be significant contributors to the risk of Achilles tendinopathy.

Polymorphisms in genes, encoding caspases (*CASP*) and nitric oxide synthases (*NOS*), have also been investigated, as these molecules were involved in pathways accompanying the tendon cell apoptosis, and their expression was elevated in tendinopathy (Nell et al., 2012). South African and Australian cohorts were genotyped for four polymorphisms: (*CASP8* rs3834129, rs1045485, *NOS3* rs1799983, and *NOS2* rs2779249). A significant association between both *CASP8* polymorphisms (rs3834129, rs1045485) and Achilles tendinopathy was found in both populations. The D/D genotype of rs3834129 was associated with tendinopathy, whilst the C allele of rs1045485 was associated with the absence of Achilles tendinopathy. *NOS3* (rs1799983), and *NOS2* (rs2779249) were not associated with Achilles tendinopathy. However, the data presented in this study showed that the control group in the Australian cohort was not in Hardy-Weinberg equilibrium (HWE), which refers to constant proportions of allele and genotype frequencies in a population, and therefore this association should be interpreted with caution. Deviations from HWE in the control group may indicate significant methodological flaws including selection bias, population stratification and genotyping errors (Namipashaki, Razaghi-Moghadam, & Ansari-Pour, 2015). Yet study was the first that investigated polymorphisms in caspase

pathways. However, recent genome-wide association study (GWAS) of almost 5,000 patients with Achilles tendinopathy failed to identify any statistically significant polymorphisms as well as replicate findings of the previously described studies of polymorphisms located in *CASP8, COL5A1, MMP3* and several other genes (Kim et al., 2017). Possibly, the failure to identify statistically significant polymorphisms and to replicate results of the previous studies was due to different characteristics of cohorts, as Kim et al. used samples and data collected from the hospital patients, who were diagnosed with Achilles tendon bursitis, tendinopathy or rupture, and Kim et al. did not control for the patients' physical activity levels. Additionally, studies used as a reference by Kim et al., utilised a candidate gene approach, and generally were underpowered and demonstrated weak statistical associations.

A tumour necrosis factor receptor 1 gene *TNFRSF1A,* which signals inflammation and apoptosis in response to the tumour necrosis factor-alpha (TNFα), was investigated as a potential gene associated with Achilles tendinopathy (Gaida et al., 2012). *TNFRSF1A* rs4149577 polymorphism was previously associated with several musculoskeletal and inflammatory diseases. However, this study by Gaida et al. was the first to investigate this polymorphism in association with Achilles tendinopathy. Another polymorphism investigated in this study was rs1049253 in the caspase-3 gene *CASP3*, which was shown to influence *CASP3* mRNA expression. Caspase-3 is involved in cellular apoptosis, including roles in chromatin condensation and DNA fragmentation (Porter & J nicke, 1999). This study also investigated the influence of the copy number variant (CNV) spanning intron 11-intron 12 in *CASP8*. CNVs are segments of DNA greater than 1kb in size that can influence phenotypes by changing gene dosage and disruption of coding sequences in DNA. The results did not show any significant association between Achilles tendinopathy, the investigated polymorphisms and CNV. This was possibly due to the study limitations, such as a relatively small sample size and a possible additional degree of error due to the rounding of copy number data into discrete calls, which, therefore, may indicate that there was no association between these CNVs and Achilles tendinopathy (Rickaby, El Khoury, Ribbans, & Raleigh, 2015) (Table 1.1). An expanded version of Table 1.1, which includes sample sizes and reported *p*-values for each of the reviewed studies, can be found in Appendix 1.

Table 1.1 Summary of genes and polymorphisms investigated in association with Achilles tendinopathy.

Gene	Polymorphisms	Product	References
TNC	GT repeats intron17 rs13321 rs2104772 rs1330363 rs2104772	Tenascin-C	(Kim et al., 2017; Mokone et al., 2005; Saunders et al., 2013)
COL5A1	rs12722 rs71746744 rs16399 rs1134170	Pro-a1 (V) chain of the type V collagen	(Abrahams et al., 2013; Brown et al., 2016; Hay et al., 2013; Kim et al., 2017; Mokone et al., 2006; Raleigh, 2009; September, 2009)
COL27A1	rs946053	Type XXVII collagen	(Saunders et al., 2013)
MIR608	rs4919510	Micro-RNA 608	(Abrahams et al., 2013; Brown et al., 2016; Kim et al., 2017)
MMP3	rs679620 rs591058 rs650108	Matrix metalloproteinase 3	(El Khoury et al., 2016; Gibbon et al., 2016; Kim et al., 2017; Raleigh, 2009)
COL12A1	rs240736 rs970547	Type XII collagen	(September et al., 2008)
COL14A1	rs4870723 rs1563392	Type XIV collagen	(September et al., 2008)
COL11A1	rs3753841 rs1676486	α1 chain of type XI collagen	(Hay et al., 2013)
COL11A2	rs1799907	α2 chain of type XI collagen	(Hay et al., 2013)
ADAMTS2	rs1054480	Tendon procollagen N-propeptidase 2	(El Khoury et al., 2013)
ADAMTS5	rs226794	Tendon procollagen N-propeptidase 5	(El Khoury et al., 2013)

Table 1.1 Summary of genes and polymorphisms investigated in association with Achilles tendinopathy (continuation).

Gene	Polymorphisms	Product	References
ADAMTS14	rs4747096	Homologue of ADAMTS2	(El Khoury et al., 2013; Kim et al., 2017)
ADAM12	rs3740199	Disintegrin and metalloproteinase 12	(El Khoury et al., 2013)
TIMP2	rs4789932	Tissue inhibitor of metalloproteinases 2	(El Khoury et al., 2013; El Khoury et al., 2016; Kim et al., 2017)
FBN2	rs331079	Fibrillin-2	(El Khoury et al., 2015; Kim et al., 2017)
ELN	rs2071307	Elastin	(El Khoury et al., 2015)
TGFB1	rs1800469	Transforming growth factor-β1	(Posthumus et al., 2010)
GDF5	rs143383	Growth/differentiation factor-5	(Posthumus et al., 2010)
IL1RN	rs2234663	IL-1β receptor antagonist	(September et al., 2011)
IL1B	rs1143627 rs16944	Interleukin-1β	(September et al., 2011)
IL-6	rs1800795	Interleukin-6	(September et al., 2011)
CASP8	rs3834129 rs1045485 CNV (intron 11, 12)	Caspase 8	(Kim et al., 2017; Nell et al., 2012)
NOS3	rs1799983	Nitric oxide synthase 3	(Nell et al., 2012)
NOS2	rs2779249	Nitric oxide synthase 2	(Nell et al., 2012)
TNFRSF1A	rs4149577	Tumour necrosis factor receptor 1	(Rickaby et al., 2015)
CASP3	rs1049253	Caspase 3	(Rickaby et al., 2015)

While these studies indicate some links between polymorphisms in genes involved in apoptosis, inflammation and Achilles tendinopathy, the majority of these studies were unable to demonstrate that these genes are probable risk factors for Achilles tendinopathy. The processes of apoptosis and inflammation contribute to the pathology of Achilles tendinopathy, however, further investigation should be undertaken to clarify the role of genetic variation, which may impact these pathological processes.

1.5 Bone stress injuries

1.5.1 Bone structure and pathology

The skeleton is a metabolically active organ which changes and adapts under mechanical stimuli by remodelling bone structure (Karlsson & Rosengren, 2012). Bone is a connective tissue with cells embedded in a mineralised matrix composed of collagen, mainly collagen type I (30 of the matrix) and inorganic salts rich in calcium and phosphates (70 of the matrix). Bone tissue exists in two forms: compact and cancellous. The compact bone constitutes the surface of the bones and consists of collagen fibres arranged in layers with embedded osteocytes. Cancellous bone is a sponge type tissue, also known as trabeculae and is found in the interior of bones. Bone modelling is a process when osteoclasts form bone tissue by producing collagen matrix, which leads to an increase in bone mass. This process occurs during the growth period of the skeleton and defines bones shapes and sizes. Bone remodelling is a lifelong renewal process of the bone, which is responsible for bone maintenance and repair (Kiuru, Pihlajam ki, & Ahovuo, 2004). The bone tissue can be remodelled by osteoclasts, which resorb matrix to mineral content, and then osteoblasts deposit new bone matrix (Figure 1.5) (Sinnatamby, 2011). The ability of the bone to remodel is essential as it allows the bone to adapt to the mechanical loads and modulate its density. Bone requires the load to develop normally, and if the load is eliminated and osteoblastic function decreases, this may lead to low diffuse osteoporosis (Kiuru et al., 2004). However, under the increased load, osteoclastic and osteoblastic functions may become unbalanced. Therefore, when there is a gap in time between the high osteoclastic activity and the high osteoblastic activity, the bone is weakened during remodelling, predisposing it to microfractures (Bennell, Matheson, Meeuwisse, & Brukner, 1999; Mattila, Niva, Kiuru, & Pihlajam ki, 2007).

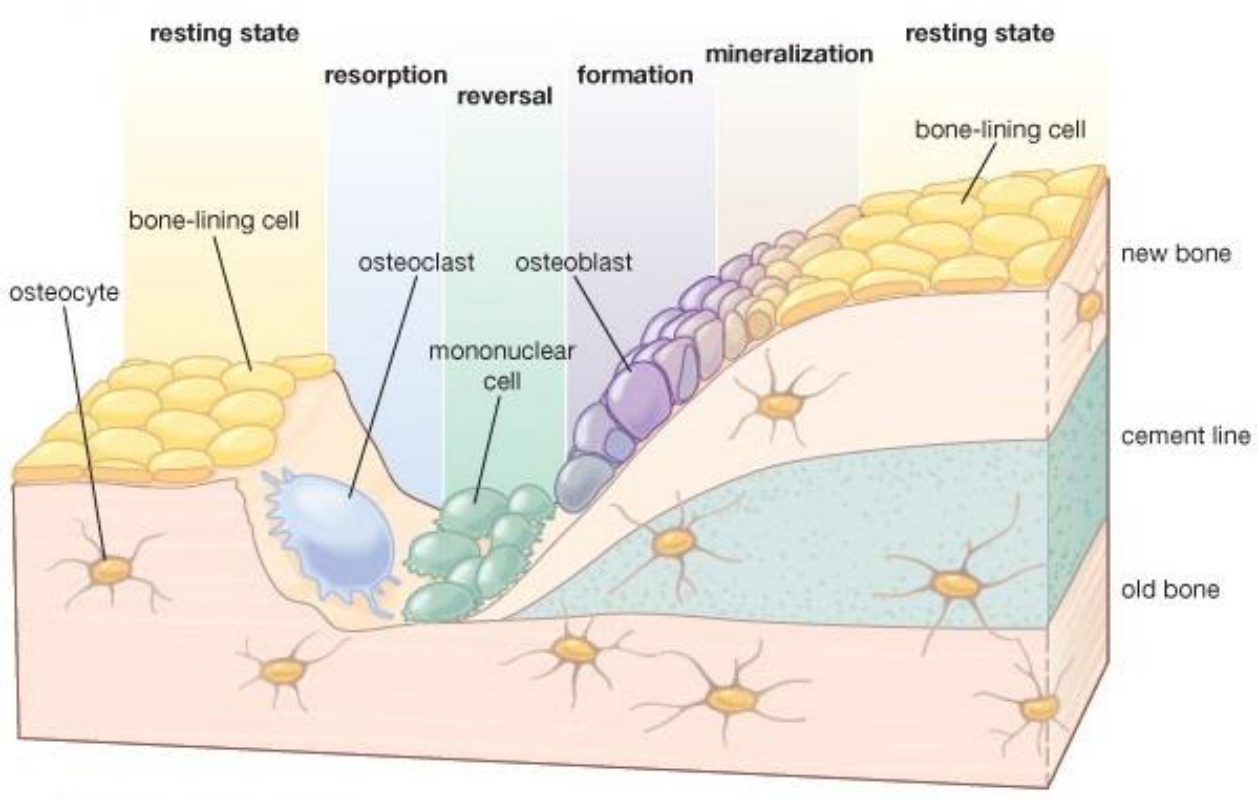


Figure 1.5 Bone remodelling process (adapted from (Encyclopaedia Britannica, 2017)).

A bone stress injury is an overuse injury attributed to the repetitive loading of the bone with vigorous weight-bearing activity, such as marching, running and jogging (Mattila et al., 2007). Initially, this injury was recognised in soldiers, and the most prevalent diagnosis was a metatarsal stress fracture. Later, with the increase of athletic participation, bone stress injuries were also noted among civilians, particularly athletes (Flinn, 2002). Stress fractures occur under two circumstances. Fatigue fractures occur when the normal bone is exposed to repeated abnormal stress, while insufficiency fractures occur in the abnormal bone under normal stress (Mattila et al., 2007). Pathologically, bone stress injuries begin with stress reactions, which can develop into stress fractures and finally complete bone fractures (Figure 1.6) (Warden, Davis, & Fredericson, 2014). Typical symptoms can be characterised by local pain, which occurs during impact loading, and worsens if the loading continues and eventually can persist after the impact loading has finished (Pollock, 2011). The most common locations of the bone stress injuries are the tibial shaft and metatarsal bones, whereas pelvis, hip, thigh and knee are less common anatomic locations (Korpelainen, Orava, Karpakka, Siira, & Hulkko, 2001b; Mattila et al., 2007). In a systematic review of

running-related injuries, medial tibial stress syndrome (MTSS) was shown to be the most prevalent bone injury (9.5) associated with running (Dias Lopes et al., 2012). A recent prospective cohort study of 933 novice runners also demonstrated that 15 of injured runners developed MTSS (Figure 1.7) (Nielsen, Ronnow, Rasmussen, & Lind, 2014).

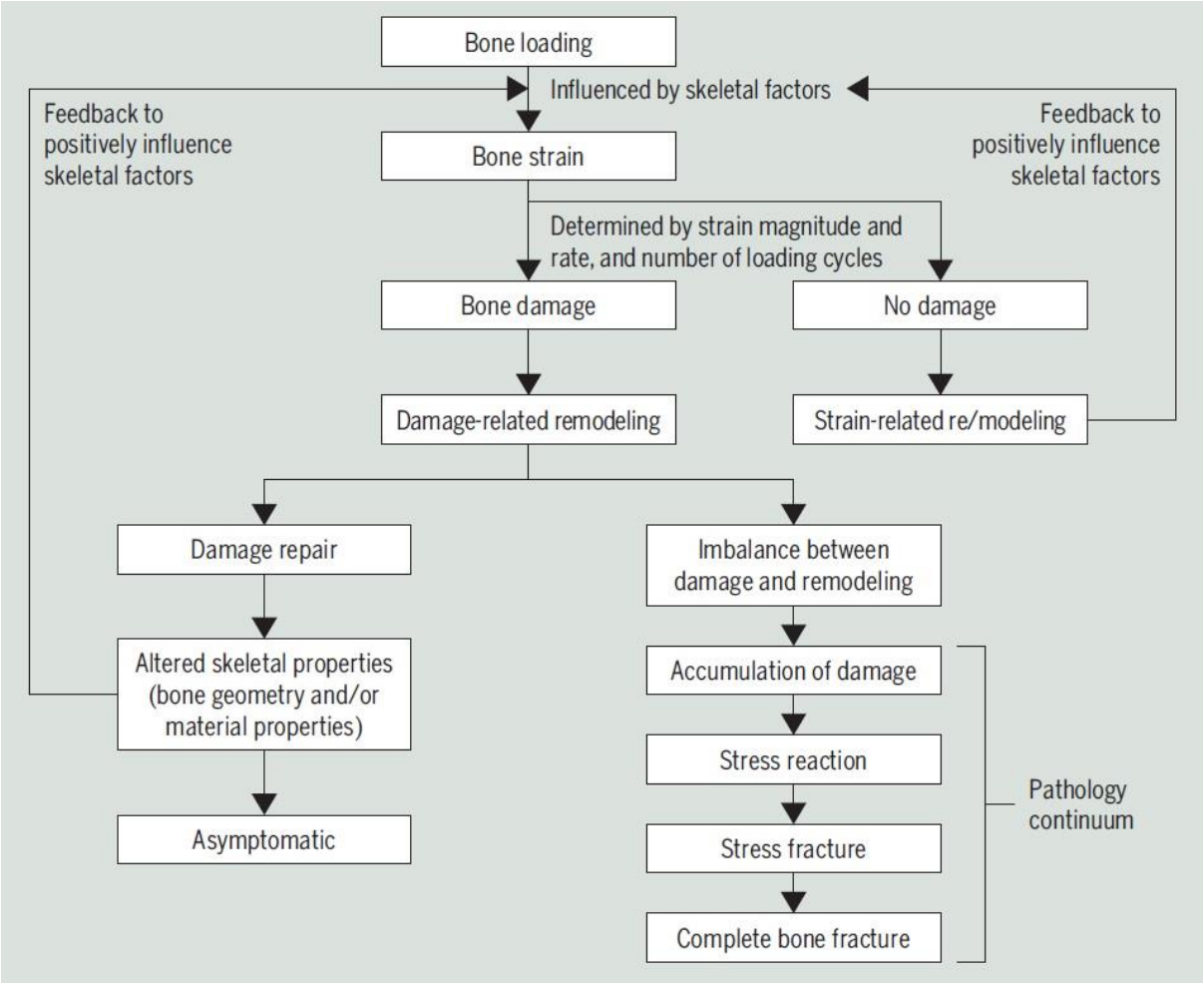


Figure 1.6 Pathophysiologic mechanisms of bone stress injury development (adapted from (Warden et al., 2014)).

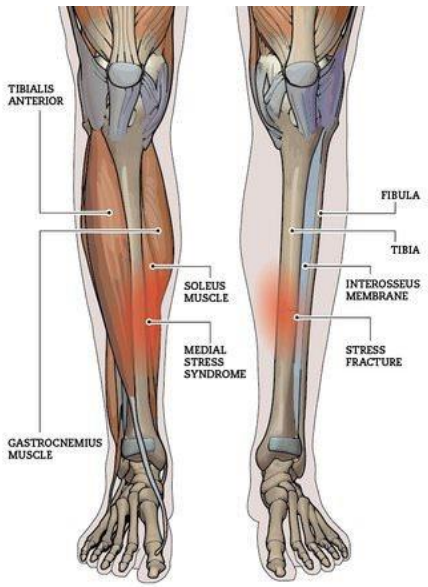


Figure 1.7 A typical pain localization with medial tibial stress syndrome (adapted from (Walsh, 2017)).

Magnetic resonance imaging (MRI) was shown to be the most accurate and reliable image assessment of bone stress injuries, including early stages of these types of injury (Beck et al., 2012). The management of bone stress injuries depends on their anatomic location and severity. Firstly, physical activity should be modified, considering the location of the injury. The damaged bone should be unloaded after the injury occurrence. Partial weight-bearing exercise can be introduced in the next several weeks, depending on injury severity. Secondly, the rehabilitation process should include muscle strengthening and a gradual return to usual physical activity (Flinn, 2002; Pollock, 2011). Prevention of bone stress injury occurrence is the priority of injury management. Bone strength can be maximised by analysing and correcting intrinsic factors and directed training to help to improve bone strength (Pollock, 2011).

Due to the complexity of the interaction between factors contributing to the development of bone stress injuries, risk factors may be divided into factors modifying the load applied to a bone, and factors influencing bone density, and hence the ability of bone to resist the load. The first group of factors includes biomechanical traits, training habits and environmental conditions. The second group of factors includes sex, age, endocrine and hormonal status, chronic disorders, physical activity levels, diet and nutrition, therapeutic agents and genetic factors (Warden et al., 2014).

1.5.2 Factors modifying the load applied to the bone

1.5.2.1 Anatomic and biomechanical characteristics

Differences in leg length showed a consistent association with stress fractures both in military and civilian populations and can be considered as a key anatomic risk factor (Bennell et al., 1999). Abnormal movement patterns in runners can also increase the risk of bone stress injuries; for instance, static alignment may influence movement patterns and hence be implicated in bone stress injury development (Warden et al., 2014). Due to variations in anatomy, different anatomic and biomechanical characteristics are associated with bone stress injuries in males compared to females. For instance, increased external rotation hip range of motion is a risk factor of MTSS in men (Newman, Witchalls, Waddington, & Adams, 2013), whereas narrow pelvis (<26cm) is associated with a greater risk of stress fractures among female marines (Winfield, Moore, Bracker, & Johnson, 1997).

1.5.2.2 Running habits and training characteristics

Bone adaptation time for increasing loads should be considered in the development of a training regimen. Repetitive mechanical load contributes to stress fracture development. The load applied to the bone is the result of the summation of external and internal forces, which are determined by training factors, physical fitness and anatomy. Several military studies reported a correlation between the levels of previous physical activity and rates of stress fractures during training (Bennell et al., 1999). For instance, poor physical conditioning measured in male Finnish conscripts was associated with an increased risk of stress fractures (V lim ki et al., 2005). Changes in the training regimen, such as rest periods, elimination of running and marching on concrete, reduction of high impact activity may reduce the risk of stress fractures (Bennell et al., 1999). Incrementing a training program too rapidly may be the central training factor, which leads to the microdamage accumulation and disruptions in bone turnover (Warden et al., 2014). In addition, a systematic review of risk factors of MTSS demonstrated that fewer years of running experience was associated with the development of the injury (Newman et al., 2013). However, the cut-off values for running experience and MTSS risk remained undetermined. Overall, the longer experience of training and physical activity appears to be protective against bone stress injuries (Warden et al., 2014). Training surface may contribute to the development of stress injuries. The ability to adapt leg stiffness to the surface makes this

risk factor very complex to study. Runners alter their leg stiffness when running on surfaces of different compliance in order to maintain a constant vertical excursion of their centre of mass. Therefore, changes in running surfaces might increase the risk of injuries as a runner must adapt their biomechanics to the new training conditions (Warden et al., 2014).

Athletic footwear and orthotics aim to attenuate shock with ground contact and to control the motion of the foot and ankle. Indeed, the choice of appropriate training shoes is essential for injury prevention. However, the majority of the studies showed a limited effect of wearing insoles and orthotics in relation to stress fracture development. However, a systematic review of MTSS showed that orthotic use was associated with increased risk of the injury (Newman et al., 2013). Military studies demonstrated reduction of overuse injuries in the foot, but not in tibia when military boots were replaced by athletic shoes during training. Further studies showed that wearing semi-rigid orthotic devices decreased the incidence of femoral stress fractures, but not overall incidence. The inconsistency of results in these studies may be explained by the interaction between sports shoes/orthotic use, the anatomy of the foot and the site of stress fracture (Bennell et al., 1999).

1.5.3 Factors affecting bone density

1.5.3.1 Physical characteristics of recreational runners with bone stress injuries below the knee

Incidence of bone stress injuries was significantly higher in female runners in several studies, and a systematic review of risk factors of MTSS reported that there was a 1.7-fold higher risk of developing MTSS for female runners than male runners in 9 analysed studies (Newman et al., 2013). Several studies of military populations showed a significantly increased incidence of bone stress injuries among female conscripts (Mattila et al., 2007; Protzman & Griffis, 1977). However, this association may be due to sex-related factors, such as hormones, menstrual cycle, bone density, diet and anatomic characteristics (Tuan, Wu, & Sennett, 2004). The term ‘female athlete triad’ was established in order to describe three major factors of bone stress injuries in female athletes: low bone mineral density (BMD), menstrual disturbances and nutritional issues. The underlying mechanism of these co-factors is a negative energy balance, developing if energy expenditure is higher than energy intake (Korsten-Reck, 2011). A cohort study of 259 physically active female adolescents showed that a cumulative risk of bone stress injuries increases as the number of triad-related

factors accumulates (Barrack et al., 2014). However, in April 2014, the International Olympic Committee published a consensus statement and introduced a new, more comprehensive and broad term – Relative Energy Deficiency in Sport (RED-S). The previous term 'female athlete triad' was exclusive to women, and did not consider some negative outcomes of low energy availability (Mountjoy et al., 2015). Some indicated that energy deficiency may also occur in male athletes, and they also may have dietary disorders, hormonal disruption and impaired bone health (Bennell, Brukner, & Malcolm, 1996a). RED-S allows athletes and coaches to take into account numerous health consequences of energy deficiency and sports clinicians can apply this model to a range of athlete clinical presentations (Mountjoy et al., 2015). For instance, low BMI (<20kg/m^2) is typical in case of RED-S; however, its association with the increased risk of bone stress injuries may also be explained by poor muscle strength, which could lead to injuries (Mattila et al., 2007).

Bone density decreases with age, and therefore, the ability of bone to endure overloading is reduced with age. Bones with lower density are more likely to accumulate microdamage leading to the stress fractures in older age. On the other hand, children and adolescents may be at risk of stress fractures due to the immature condition of their bones. Hormonal status and training loads may be important co-factors of age as a risk factor (Bennell et al., 1999).

Body size and BMI were associated with the development of bone stress injuries in several studies. Increased BMI was associated with the development of MTSS (Newman et al., 2013). A case-control study of tibial stress injury (TSI) showed that both men and women with TSI had significantly higher body fat and lower lean mass than uninjured matched controls (Beck, Rudolph, Matheson, Bergman, & Norling, 2015).

1.5.3.2 Lifestyle habits

Smoking and alcohol consumption may have an adverse effect on bone turnover by limiting calcium absorption (Pollock, 2011). Numerous studies demonstrated an inverse relationship between smoking and bone mineral density. A meta-analysis of smoking and stress fracture risk showed that smoking was associated with reduced bone mineral density in men and post-menopausal women. Even though both 'having ever smoked' and 'current smoking' contribute to the increased risk of stress fractures, current smoking had higher risk ratios of stress fractures than ever-smoking (Kanis et al., 2005). A prospective study of 3,758 female

US army recruits showed that the history of smoking and current smoking increase risk of stress fractures among young female recruits, moreover, the relative risk increased with more years of smoking and more packs of cigarettes per day (Lappe, Stegman, & Recker, 2001).

Long-term excessive alcohol consumption was associated with low bone mass in both males and females. The same study of the female conscripts also reported that alcohol consumption of 10 or more alcoholic drinks per week was a risk factor of stress fractures (Lappe et al., 2001). Unfortunately, it was hard to ascertain alcohol consumption as an independent risk factor because many participants who consume alcohol also smoked in this study.

1.5.3.3 Diet and nutrition affecting bone health

Eating disorders are common among athletes when individual low weight and lean body are a pre-condition for high performance. Anorexia athletica is an eating disorder, which along with anorexia nervosa may lead to disrupted endocrine status and following osteoporosis and proneness to bone stress injuries (Korsten-Reck, 2011).

Calcium is an essential component of bone mineralisation, and vitamin D is an essential factor of calcium absorption. The vitamin D metabolite 1,25-dihydroxyvitamin D (1,25(OH)2D) is the most physiologically active metabolite. It acts through the vitamin D receptor directly, to increase intestinal calcium absorption and to enhance renal calcium reabsorption. It is also known to decrease parathyroid hormone secretion and enhance the differentiation of both osteoblast and osteoclast precursors. If dietary calcium is inadequate, vitamin D causes osteoclasts to mature and resorb calcium from the bone. One study demonstrated that lowered serum $1,25(OH)_2D$ levels were a significant risk factor for stress fractures among a military population of Finnish conscripts (Ruohola et al., 2006). A prospective study of young female runners showed that women whose calcium daily intake was less than 800mg had a nearly 6 times higher incidence rate of bone stress injuries when compared to women who consumed more than 1,500 mg of calcium per day (Nieves et al., 2010). Calcium and vitamin D supplementation was also demonstrated to be an efficient means of preventing bone stress injuries among female navy recruits (Lappe et al., 2008).

1.5.3.4 Hormonal status

Female athletes have a higher prevalence of menstrual disturbances when compared to the general female population (Torstveit & Sundgot-Borgen, 2005). These disturbances include delayed menarche, anovulation, oligomenorrhoea and amenorrhoea. Higher incidence of bone stress injuries in female athletes may be explained by low oestrogen levels. Oestrogen deficiency accelerates bone remodelling and causes increased calcium excretion leading to decreased bone density. The combination of these circumstances may contribute to the development of bone stress injuries. Regarding post-menopausal women, falling oestrogen levels also lead to similar consequences and also predispose older women to stress fractures and osteoporosis (Pollock, 2011). Although many studies have investigated the effect of the oral contraceptive pill on the development of bone stress injuries, the obtained results were contradictory, and the oral contraceptive pill could not be suggested as an effective preventive therapy against bone stress injuries (Bennell et al., 1999). Interestingly, other hormones related to calcium metabolism failed to show any association with the incidence of bone stress injuries, possibly due to poor sampling and measurement procedures in the studies (Bennell et al., 1999). Hormonal status is an important contributor to bone health and should be considered as a risk factor of bone stress injuries.

1.5.3.5 Medications affecting bone structure

Glucocorticoids (a class of steroid hormones) induce osteoblast apoptosis and increase osteoclast activity, these two effects leading to disruptions in bone turnover and weakened bones (Rehman & Lane, 2003). Corticosteroids have an adverse effect on bone density, causing an imbalance between osteoclast and osteoblast activities. Corticosteroid use is a significant risk factor for osteoporosis, which is included in the clinical guidelines for the assessment of osteoporosis. A meta-analysis of corticosteroid use in seven prospectively studied cohorts showed that prior corticosteroid use increased the risk of stress fractures and should be considered as an independent factor for bone mineral density and the previous fragility fracture (Kanis et al., 2004).

Patients with active epilepsy require regular medication, and therefore, in addition to the benefits of the treatment, they are exposed to the side effects of anti-epileptic drugs. The prevalence of bone disorders among patients chronically treated with anti-epileptic drugs was reported at 50 (Petty, O'Brien, & Wark, 2007). Multiple biochemical abnormalities take

place when patients take anti-epileptic drugs: reduced serum calcium, phosphate and vitamin D, increased concentration of the parathyroid hormone. Parathyroid hormone regulates calcium metabolism. When calcium concentration is decreased, the parathyroid hormone triggers bone breakdown and resorption (Pack, 2003). The mechanism of the anti-epileptic drug's impact on bone remains controversial, the most popular theory being that anti-epileptic drugs induce cytochrome p450 enzymes, which cause increased vitamin D degradation. However, some studies showed hypocalcaemia and low bone density in patients without vitamin D deficiency (Petty et al., 2007).

Cancer therapies, including radiotherapy and chemotherapy, can directly or indirectly damage bone. Several chemotherapy agents may have a direct effect on bone metabolism. For instance, methotrexate reduces osteoblast production by inhibiting DNA synthesis and simultaneously increases osteoclast production. Overall, the main side effect of chemotherapy is decreased bone formation and subsequent bone mass loss. Likewise, chemotherapy, radiation causes significant bone loss (Michaud & Goodin, 2006).

Bisphosphonates are stable pyrophosphate analogues which bind bone mineral. Bisphosphonates are prescribed to reverse bone loss and slow or stop the natural process that dissolves bone tissue, resulting in maintained or increased bone density and strength. These are prescribed for osteoporosis and in some cancer treatments (Armamento-Villareal et al., 2009; Michaud & Goodin, 2006). However, the detrimental effects of long-term bisphosphonate treatment were also found. The prolonged suppression of resorption by continuous use of bisphosphonates indirectly inhibits bone formation, which may lead to the decreased ability to repair micro-damage or rejuvenate old bone (Armamento-Villareal et al., 2009).

1.5.3.6 Chronic disorders

Several chronic disorders have been shown to be associated with an increased incidence of bone stress injuries, often due to an inherent occurrence of low bone density.

Cerebral palsy is defined as a non-progressive disturbance of the developing foetal or infant brain that results in movement and posture disorders that cause activity limitations (Rosenbaum et al., 2007). Cerebral palsy is a common chronic motor disability, with 3.6 per 1,000 children prevalence (Esen, Demirel, Güven, Değerliyurt, & Köse, 2011). Many children

and adults with cerebral palsy have diminished bone mineral density, which results in bone fractures and therefore impaired physical function and poor quality of life (Houlihan & Stevenson, 2009).

Cystic fibrosis is an inherited, chronic, progressive and fatal disease, caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Cystic fibrosis affects the lungs, liver, pancreas and intestines (Li et al., 2014). Although this disease doesn't affect bone directly, decreased bone mineral density is demonstrated in patients with cystic fibrosis. Robertson et al. reported 5.5 incidence of bone fractures among cystic fibrosis patients. Reduced bone density in cystic fibrosis patients is thought to be associated with nutritional deficiency and use of bisphosphonate medication (Robertson & Macdonald, 2010).

Rheumatoid arthritis is an autoimmune disease which causes irreversible joint damage and functional disability (Fechtenbaum, Nam, & Emery, 2014). Patients with rheumatoid arthritis have a greater risk of osteoporosis than the general population due to their impaired walking ability, chronic inflammation and glucocorticoid use. A prospective study of fracture incidence in rheumatoid arthritis patients reported a significantly higher risk of fractures than in previous retrospective studies, and the majority of occurred fractures caused a gait disturbance in patients (Nampei et al., 2008).

1.5.3.7 Genetic markers associated with bone stress injuries

Several signs of the existence of genetic variants have been mentioned in the literature: the presence of stress fractures in monozygotic twins (Singer, Ben-Yehuda, Ben-Ezra, & altzman, 1990) and high recurrence rates of bone stress injuries in track-and-field athletes (Bennell et al., 1996b). There is minimal knowledge of the genetic contribution to the development of bone stress injuries and findings are usually based on studies of genetic markers of osteoporosis and exercise. Osteoporosis is a multifactorial disorder in which loss of bone strength leads to fragility fractures. A large genetic study of 2018 male patients with osteoporosis searched for polymorphisms in the genes that play an important role in bone structure and remodelling and could be associated with bone mineral density. This study found 11 polymorphisms in 10 genes that showed a strong association with vertebral bone mineral density (muda et al., 2011). In addition to calcium and vitamin D supplementation and oestrogen levels, around 30 genetic markers, involving a vitamin D receptor gene

(*VDR*), bone morphogenetic protein (*BMP2*) genes showed association with the development of osteoporosis (Raisz, 2005; Styrkarsdottir et al., 2003).

The Wnt signalling pathway has extensive functions and regulates processes related to cell growth, differentiation, function and cell death during embryonic development as well as adult life. This pathway plays an essential role in bone formation, and the decreased functioning of the Wnt pathway results in osteopenia or osteoporosis (Piters, Boudin, & Van Hul, 2008) Polymorphisms in genes encoding a protein involved in this pathway (*LRP5*) or a protein-antagonist of this pathway (*SOST*), were shown to be associated with significant increase or decrease of bone mineral density (Piters et al., 2008).

The most extensively studied gene in relation to stress fractures is the vitamin D receptor (*VDR*) gene. The vitamin D receptor plays a vital role in calcium metabolism as it is located in various tissues, including intestine and bone, and mediates the effects of vitamin D (Kehoe & Montgomery, 2006). Several polymorphisms in *VDR* were found to be associated with either increased or decreased bone mineral density (Ferrari, Rizzoli, Slosman, & Bonjour, 1998; Kehoe & Montgomery, 2006). Another investigated genetic marker is a calcitonin receptor gene (*CALCR*). Calcitonin is a hormone involved in osteoclast activity. A case-control study of 203 uninjured soldiers and 182 soldiers with stress fractures showed that several SNPs in *CALCR* (rs12154667, rs1548456) and *VDR* (rs4328262) genes were associated with stress fractures (Yanovich et al., 2012).

Genetic analysis of 518 DNA samples obtained from elite athletes with radiologically confirmed stress fractures showed that several polymorphisms in genes within the RANK/RANKL/OPG signalling pathway were associated with stress fracture susceptibility. The RANK/RANKL/OPG signalling pathway plays an important role in the regulation of bone remodelling and bone adaptation (Varley et al., 2015). Another candidate gene studied in the same elite athletes, and a cohort of conscripts is *P2X7* that encodes a highly purinergic P2X7 receptor. This receptor is expressed by osteoblasts and osteoclasts and involved in cellular responses to stress. Identified significant SNPs (rs3751143 and rs1718119) have also been shown to be associated with bone phenotypes in other studies (Varley et al., 2016) (Table 1.2). An expanded version of Table 1.2 can be found in Appendix 2.

Table 1.2 Summary of genes and polymorphisms investigated in association with bone stress injuries.

Gene	Polymorphisms	Product	References
VDR	rs4328262	Vitamin D receptor	(Ferrari et al., 1998; Kehoe & Montgomery, 2006; Yanovich et al., 2012)
CALCR	rs2051748 rs12154667 rs1548456	Calcitonin receptor	(Yanovich et al., 2012; muda et al., 2011)
BMP2	Not specified	Bone morphogenetic protein 2	(Raisz, 2005; Styrkarsdottir et al., 2003)
LRP5	Not specified	Low-density lipoprotein receptor 5	(Piters et al., 2008)
SOST	rs1877632	Sclerostin	(Piters et al., 2008; muda et al., 2011)
RANK	rs3018362	Receptor activator of nuclear factor-KB	(Varley et al., 2015)
RANKL	rs1021188	Receptor activator of nuclear factor-KB ligand	(Varley et al., 2015)
P2X7	rs3751143 rs1718119	Purinergic receptor	(Varley et al., 2016)

1.6 Genome-Wide Association Studies

The development of high-density genetic methods has led to a shift towards a genome-wide association study (GWAS) approach based on the linkage disequilibrium (LD) principle (Hartl, Clark, & Clark, 1997), specifically, the non-random association of genetic variants at different genomic loci in a population. Loci are said to be in LD when the frequency of association of the different genetic variants is higher or lower than what would be expected if they were independent and associated randomly. This technology allows researchers to identify multiple genetic markers, from hundreds of thousands-to-millions, simultaneously across the whole genome and avoid a bias of preselection of potentially important genes as in the candidate gene approach. However, genome-wide association studies have several limitations. Specifically, GWAS require a large sample size to achieve adequate statistical power and provide statistically significant results. The sample size of GWAS depends on many factors: disease prevalence, disease allele frequency, LD, effect size and inheritance models (Hong & Park, 2012). As GWAS evaluate hundreds of thousands of SNPs, this leads to multiple comparisons and increases the risk of false-positive results. Therefore, the *p*-value threshold must be adjusted, usually through the employment of a Bonferroni correction (standard *p*-value of 0.05 divided by the number of analysed SNPs (Abdi, 2007). However, with a large sample size, it may be difficult to control for homogeneity of the studied groups and consistent and precise diagnoses. Finally, the effect size of the association is important for the application of the findings in clinical practice. According to Klein et al., only 6.8 of GWAS reported Odds Ratio (OR) of more than 3.0 at a *p*-value less than 10^{-5} (Klein, Lohmann, & iegler, 2012).

In relation to exercise and health, several consortiums have been established to collect data and samples in order to reach a sufficient sample size. The HERITAGE family study (HEalth, RIsk factors, exercise Training and GEnetics) was organised by five universities to investigate the role of genetic variants in cardiovascular, hormonal and metabolic responses to aerobic exercise training (Bouchard et al., 1995). In addition, the Athlome Project Consortium has been established in order to study genetic and phenotypic data of athletes in adaptation to exercise and on exercise-related injuries (Pitsiladis et al., 2016). This Consortium investigates data collected from cohorts of athletes across many countries such as the UK, Japan, Eastern European countries, East African countries, and continues to expand their collaborative efforts and grow numbers of participants in each of the project. A recent publication of the GWAS on Achilles tendinopathy and Anterior Cruciate Ligament

(ACL) injuries presented results from the Genetic Epidemiology Research on Adult health and ageing (GERA) cohort of over 100,000 people (Kim et al., 2017). These examples of collaborative efforts and large-scale studies, which aimed to identify genetic variants in association with exercise and exercise-related injuries indicate the validity and great potential of the GWAS approach in this field.

1.7 Conclusion

Physically active people, particularly athletes, are more likely to develop overuse injuries due to the biology of these types of injuries. The consequences of overuse injuries have a detrimental effect on the individual's health, quality of life and sports performance. The main priority in managing lower limb injuries in tendon and bone is to prevent these injuries. Treatment is expensive and time-consuming and, according to the relatively high injury recurrence rates, not always efficient. In order to reduce injury incidence, contributing risk factors should be taken into account and corrected if possible. This literature review of overuse injuries demonstrates several gaps in the current knowledge base, particularly in regard to genetic risk factors. The reviewed risk factors of Achilles tendinopathy and bone stress injuries represent a complex interrelationship between intrinsic and extrinsic contributors for each type of injury. Contradictory outcomes between some studies and low repeatability of the results necessitate comprehensive investigations of the risk factors, their significance and interactions. Prospective studies have demonstrated reliability with respect to training and biomechanical risk factors, as they allow the collection of more precise data by following the participants of the studies. Although genetic markers can be studied retrospectively, the data on the genetics of injuries is limited to several case-control studies on relatively small cohorts. This thesis will explore the genetics of overuse injuries in a cohort of recreational runners to attempt to provide a better understanding of the genetic contribution to the development of overuse injuries and the biological processes and pathophysiology underlying their development.

1.8 Project overview

1.8.1 Purpose of the study

The purpose of the study is to identify training and lifestyle-related characteristics of physically active people, genetic polymorphisms associated with Achilles tendinopathy and bone stress injuries, and to systemise and make these data available for the preventive strategy against running-related injuries. This study will use both epidemiological and GWAS approaches to investigate training and lifestyle-related factors associated with running-related injuries and identify genetic polymorphisms associated with the incidence of Achilles tendinopathy and bone stress injuries in lower extremities. Due to the high popularity of running in Australia and a relatively high incidence of overuse injuries while running, recreational runners were selected as the study group. An online questionnaire approach was selected as a method to collect phenotypic data from Australian recreational runners, and to provide a platform for the selection of appropriate candidates for the GWAS. The main hypothesis of this project is that certain discovered genetic polymorphisms will be associated with either increased or decreased risk of Achilles tendinopathy and/or bone stress injuries.

Specific objectives of the project:

1. To collect and analyse health, lifestyle and training data from Australian recreational runners, and describe the training and lifestyle habits of the runners.
2. To identify and report frequencies of the running-related injuries and investigate risk factors contributing to the risk of these injuries in recreational runners.
3. To investigate which training and lifestyle habits may be risk factors of Achilles tendinopathy in recreational runners.
4. To investigate which training and lifestyle habits may contribute to the risk of bone stress injuries in recreational runners.
5. To collect and analyse DNA samples of the runners using a GWAS approach and investigate which genetic markers are associated with either increased or decreased risk of Achilles tendinopathy. Following imputation of additional genotypes will allow replication of previously identified associations from the studies which utilised a candidate gene approach.
6. To collect and analyse DNA samples of the runners using a GWAS approach and investigate which genetic markers are associated with either increased or decreased

risk of bone stress injuries. Following imputation of additional genotypes will allow replication of previously identified associations from the studies which utilised a candidate gene approach.

1.8.2 Significance of the study

This study will be the largest cross-sectional study conducted of Australian recreational runners. Collected data will allow the analysis of training patterns and injury frequencies in a large number of physically active Australians. This knowledge may help to promote physical activity among less active groups of the population as well as contribute to running injury prevention programs. The use of GWAS as a discovery-based and unbiased approach will allow investigation of multiple genetic polymorphisms simultaneously. Findings of the genetic arm of the study may contribute to a better understanding of the molecular pathways and mechanisms underlying the development of running-related injuries.

1.8.3 Sample size estimation and power calculations for GWAS

The sample size is a crucial aspect of study design that researchers can control and change in order to reach the expected power of the study and the significance level of GWAS. As a GWAS analyses hundreds of thousands of genetic variants, they require a large sample size in order to reach estimated significance levels. Initially, statistical power calculations for the project were determined as the GWAS significance level of $p<5x10^{-8}$ and a minimum threshold of the expected power at 80 (Cohen, 1988). Additional parameters required for power calculations were disease prevalence, minor allele frequency (MAF) and relative risk (RR). Thus, disease prevalence was set up at an average of 10 for each type of injury (Section 1.4.2 and Section 1.5.1). The MAFs of previously investigated polymorphisms, which were described in the literature review varied between 0.05 and 0.5 (Section 1.4.4 and Section 1.5.3) and six values were included in the power calculations: 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5. Finally, three values of RR were employed in the power calculations: 1.3, 1.5 and 2.

In order to investigate the required sample size with variation in MAFs and RR, an online power calculator for GWAS (Power-Calculator, 2015; Skol, Scott, Abecasis, & Boehnke, 2006) was utilised. The results of these calculations are displayed in Figure 1.8. If the RR is high (RR 2) and the MAF is high (MAF 0.5), then the sample size may be as small as 150

cases and 150 controls, whereas for interrogation of the SNPs with the lowest MAF of 0.05 the sample size must include 1,000 cases and 1,000 controls. Therefore, when the RR is high, and the sample size is large, statistical analysis would permit an investigation of SNPs with low MAFs. When the RR is 1.5, then a sample size of 800 cases and 800 controls will be required to interrogate SNPs with MAFs of at least 0.3. Less common SNPs will require a sample size of over 1,200 cases and 1,200 controls. Finally, if the RR is 1.3, then even the most common SNPs with MAF of 0.5 will require a sample size of 20,000 cases and an equal number of controls. These calculations demonstrated that at least 800 cases and 800 controls would be a sufficient sample size for investigation of the identified common SNPs at the RR of 1.5 and the required significance level.

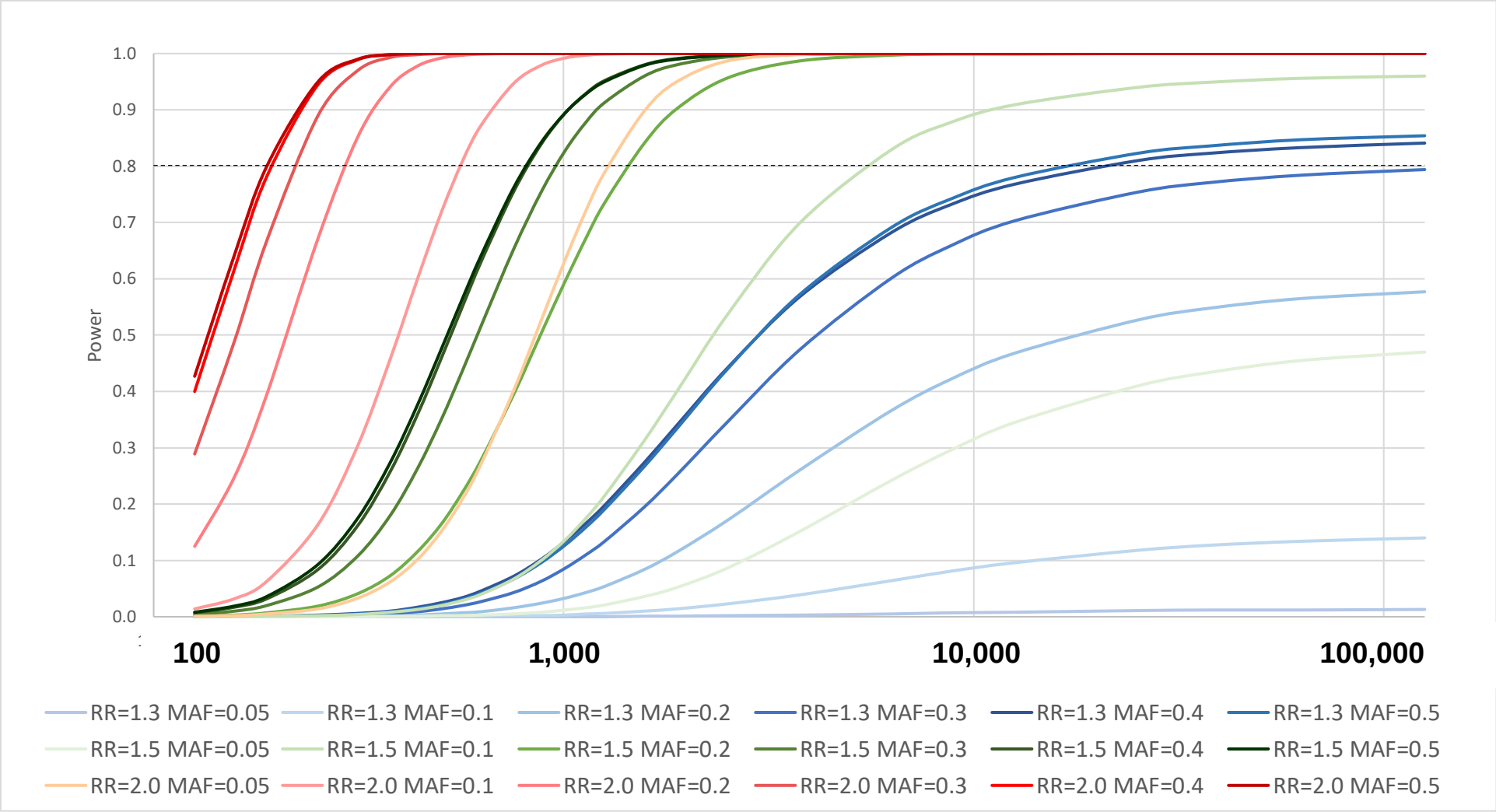


Figure 1.8 Required sample size to reach power at 80% with combinations of ranged minor allele frequencies (MAFs) and relative risk (RR).

Blue, green and red colour spectrums refer to 1.3, 1.5 and 2 values of relative risk, respectively. Each colour spectrum reflects the range of MAFs from 0.05 to 0.5. The required power level of 80 is shown by the black horizontal dashed line.

2. Chapter Two – Participant recruitment and description of health, lifestyle and training habits of Australian recreational runners

Addendum

Contributions to Chapter 2:

Mariia Kozlovskaia:

- Recruitment of participants
- Data collection
- Database management
- Data quality control
- Statistical data analyses
- Author of the chapter

Nicole Vlahovich:

- Online survey design and development
- Ethics applications
- Recruitment of participants
- Database management
- Editing of the chapter

Evelyne Rathbone:

- Assistance with statistical analyses

Silvia Manzanero:

- Assistance with recruitment of participants
- Assessment of recruitment strategies

David Hughes:

- Assistance with recruitment of participants
- Editing of the chapter

Publication outcomes of this chapter:

1. Kozlovskaia M., Vlahovich N., Rathbone E., Manzanero S., Keogh J., Hughes D.C., 2017. A profile of health, lifestyle and training habits of 4720 Australian recreational runners – the case for promoting recreational running for health benefits. Health Promotion Journal of Australia.
2. Manzanero S., Kozlovskaia M., Vlahovich N., Hughes D.C., 2018. Recruitment and Participation of Recreational Runners in a Large Epidemiological and Genetic Research Study: Retrospective Data Analysis. JMIR Research Protocols.

2.1 Introduction

Running is a popular recreational pastime. In Australia, the rate of participation in running increased from 4.3 of the population in 2005–06, to 7.4 in 2013–14 (Australian Bureau of Statistics, 2015). Recent data have indicated that current participation levels in athletics, including running, in the Australian population may be as high as 15.8 (Australian Sports Commission, 2016). Running provides a low-cost option for increasing physical activity, without the restrictions of specific equipment or costs of sports club membership. Furthering the understanding of the running habits and wider health characteristics of male and female recreational runners may assist in the development of sex-specific messaging to promote the health benefits of recreational running as a form of physical activity.

This study aimed to recruit a large sample size of approximately 10,000 participants in order to conduct case-control studies on the obtained phenotypic and genetic data and identify training-related factors and genetic variants associated with two most common running-related injuries – Achilles tendon injuries and bone stress injuries (Dias Lopes et al., 2012). The design of the study required gathering high volume data from Australian recreational runners across the country and subsequently, a remote collection of biological material from prospective participants for the genetic arm of the study.

The purpose of the descriptive part of this study was firstly, to describe the health and lifestyle characteristics of Australian recreational runners and compare body mass index (BMI) to the general Australian population, and secondly, to examine the similarities and differences in training habits of male and female runners. It was hypothesised that recreational runners would display characteristics of a healthy lifestyle, including participation in physical activity, maintaining a healthy BMI and having low rates of chronic disease and smoking.

2.2 Methods

2.2.1 Online Questionnaire

An online survey was developed utilising the SurveyGizmo Platform (Boulder, CO, USA), and comprised 185 questions with an expected average time of 30 minutes for completion. This survey contained questions covering a number of topics in the following order: physical characteristics, ethnic background, running habits, injury history (e.g. injured/uninjured in prior two years and details about injuries), history of chronic conditions, dietary habits (e.g. nutritional requirements and dietary supplements), and female health (for female participants e.g. menstrual cycle) (Appendix 5). The majority of the health, lifestyle and training-related questions were closed-ended and were either dichotomous or offered multiple choice answers. However, several questions were open-ended, which, for example, allowed respondents to provide diagnoses of injuries different from Achilles tendon injuries or bone stress injuries, dietary preferences, lost weight in kilograms.

A test-retest reliability study demonstrated that this self-reported, retrospective questionnaire is able to provide stable and reliable data and meets the recruitment needs for the project (Domaschenz, Vlahovich, Keogh, Compton, & Hughes, 2015). This survey also enabled the selection of eligible participants and subsequent sample collection for the GWAS. Several selection criteria for a GWAS have been indicated based on the literature review and further discussed in chapters 4 and 5:

- 18-50 years of age,
- reported background of at least 75 of Caucasian European or Mediterranean
- current non-smoker with at least five years after giving up,
- have no history of chronic conditions that could affect the musculoskeletal system (osteoarthritis, osteoporosis, rheumatoid arthritis, chronic renal failure),
- no history of chemotherapy.

An essential requirement for eligibility was acceptance of the following survey items: ‘I give permission, if I am eligible, to be contacted in the future for related research’ and ‘I give permission, if I am eligible, to be contacted in the future to provide a saliva sample for genetic related analysis’. The study was approved by the Bond University Human Research Ethics Committee (approval RO1688B) (Appendix 4).

2.2.2 Recruitment strategies

An overview of the project on the AIS website and SurveyGizmo platform described its aim to recruit Australian recreational runners being age over 18 years and recreational running practice of more than 15 km per week. Hence participants who identified themselves as recreational runners, self-selected for inclusion into the study with consideration of these two inclusion criteria – minimal age of 18 years and minimal weekly running distance of 15 km.

In order to simplify an online search for the study and its promotion, it was given a short title, the 'AIS Running Injury Study'. The study was actively promoted using websites of the involved research institutions, social media, running organisations and events (Appendices 6 and 7). Some of the recruitment activities were free of charge, and while others were paid for (Table 2.1). Free strategies comprised partnerships with running clubs and organisations such as *parkrun*, the establishment of a study Facebook page, voluntary referrals by runners and medical professionals, and collaborations with sports-related businesses to provide respondents with online discount codes upon survey completion. Methods that incurred a cost comprised advertising on websites, including Facebook and running-related websites, and researchers' attendance of ten running-related events, including a three-day *CityFit Expo*.

The enrolment period spanned September 2014 until October 2016. All respondents provided informed consent to participate and provided personal data after accepting the conditions of the study on the first page of the online questionnaire (Appendix 5). The last question of the survey 'How did you hear about this research?' was included in the survey 16 months after the commencement of the recruitment. This question aimed to identify the best-incorporated recruitment approach.

Table 2.1 Summary of recruitment strategies incorporated in the project with specified expenses.

Strategy	Methods and channels	Outcomes (other than recruitment)	Expenses
Facebook	Group page with regular posts Paid advertisements Posts in other pages	Page followers & post sharing	Advertising fees
Other social media	Twitter Instagram Newsletters		None
Online media	Relevant articles Radio interview Press interview	Presence in webpages Podcast Blog post	None
Running events	Flyers Emails to event participants Presence in race results emails Contact with running-related businesses	Further promotions	Transport & accommodation
CityFit Expo	Flyers Presence in Expo social media Contact with running-related businesses	Further promotions	Printed materials, transport & accommodation, stand booking
Parkrun	Presence at events Newsletters Facebook		None

Table 2.1. Summary of recruitment strategies incorporated in the project with specified expenses (continuation).

Strategy	Methods and channels	Outcomes (other than recruitment)	Expenses
Referrals (personal and professional)	Emails to previous survey participants Sports health professionals Word of mouth	Facebook posts Advice to patients	None
AIS	Website Social media		None
Email	Running events Running clubs Running-related businesses Fitness business, personal trainers, running coaches Triathlon and athletics state organisations	Mentions in newsletters Invitations to events Facebook, other social media Referrals	None
Incentives	Discount promo codes to participants Competitions		None

2.2.3 Statistical data analyses

Statistical analyses were conducted in IBM SPSS Statistics version 24 (SPSS, Inc.). Participants' BMI was calculated from the responses to weight (kg) and height (cm) and categorised as very underweight (<16 kg/m^2), underweight (16-18.5 kg/m^2), normal (18.5 to <25 kg/m^2), overweight (25 to <30 kg/m^2), moderately obese (30 to <35 kg/m^2), severely obese (35 to <40 kg/m^2) and very severely obese (≥40 kg/m^2).

All numerical variables were checked for normality using Kolmogorov-Smirnov and Q-Q plots, which indicated that the data were not normally distributed. Median, minimum and

maximum values and interquartile range (IQR) were calculated for physical characteristics (age, height, weight, BMI) and independently presented for the entire cohort and male and female subgroups. Mann-Whitney U test was performed to compare the distributions of these continuous variables between male and female subgroups and showed a significant difference (p<0.001) across all four variables. There were no missing data for these variables. Logarithmic transformation was attempted to correct skewness, however, the majority of the variables remained skewed after the transformation. Therefore, BMI and age were categorised, and subsequent analyses were performed using only categorical variables.

Categorical variables describing running habits and health conditions were summarised using counts and percentages. Statistical comparisons of recruitment strategies and male and female subgroups for the descriptive analysis were implemented for all categorical variables using the chi-squared (χ^2) test. This test was used to investigate whether distributions of categorical variables differ from one another. Not available (NA) data were presented for each categorical variable and comprised less than 0.5 . BMI data of recreational runners were compared to published health data from the general Australian population (18-74 years of age), which had been collected in the same way.

Multiple logistic regression was performed to assess the effect of running experience on the reported clinically significant weight loss (≥5kg), after adjusting for sex, age, BMI, participation in other sports, smoking history and injury history. Whilst weekly running distance and race pace were both independently strongly associated (p<0.001) with reported weight loss in univariable analyses, they were not used as predictors in the model due to strong association with running experience and sex, respectively. To avoid multicollinearity, only predictors of interest that were not strongly associated with each other were selected in the model. Results of the multivariable analysis are presented as adjusted odds ratios with 95 confidence intervals and p-values. A Hosmer and Lemeshow test indicated that the model fit was acceptable (χ_8^2 11.69, p 0.17). Statistical significance was set at p<0.05.

2.3 Results – Recruitment of Australian recreational runners

2.3.1 Implemented recruitment strategies

During 25 months of recruitment, 9,069 participants started entering data into the survey, however, only 5,250 complete responses were received. Recruitment progress and related recruitment activities are displayed in Figure 2.1. From February 2016 until October 2017, regular paid Facebook advertisement were employed as part of the recruitment strategy. In addition, researcher's attendance at multiple running events in Canberra, Sydney and Brisbane may be reflected by peaks on this recruitment graph.

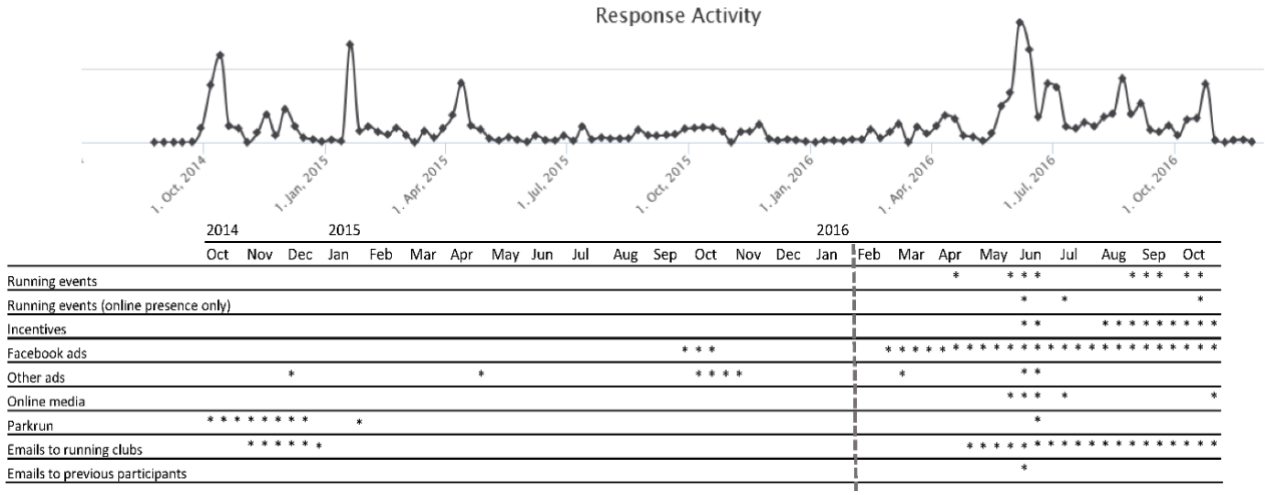


Figure 2.1 Response activity timeline and recruitment strategies implemented during 25 months.

A question about recruitment strategies was introduced to the survey only 16 months after the commencement of the project. Approximately a half (*n* 2,760) of completed responses contained data about the recruitment strategy, which led a respondent to the survey (Table 2.2). Facebook was the most popular source referenced by over a third of the respondents. Running events and *parkrun* together helped to recruit a similar number of runners as a Facebook advertisement. Almost 500 runners referred other media resources. The remaining strategies, including referrals and the AIS website and social media accounts, were mentioned as a source by approximately 5 of runners each.

Table 2.2 Summary of utilised recruitment strategies reported by participants.

Recruitment strategy	Participants	
	n	**%**
Facebook	979	35.5
Running events	618	22.4
Parkrun	368	13.3
Other social media	324	11.7
Online media (articles and interviews)	161	5.8
Referrals (personal and professional)	144	5.2
AIS	133	4.8
Other	33	1.2
Total	2760	100.0

When participants' characteristics were statistically compared by recruitment strategy, it was shown that age, sex and weekly running distance were associated with different recruitment methods (Table 2.3). Facebook recruitment showed a significant association with female participation, whereas online media appeared to recruit more male participants (p<0.001). There were significantly more runners aged 35-44 years recruited through Facebook than runners of other age groups. Older runners were more typically recruited through other social media resources and *parkrun*, whereas younger runners aged between 18 and 24 were more likely to proceed to the survey through the AIS website (p<0.001).

Participants who ran less than 20 km per week were more highly represented in the Facebook and *parkrun* categories. In contrast, those who run 40 km per week or more showed the least presence in the *parkrun* category. As the study was explicitly focused on the injury, this factor was also tested for association with recruitment methods. Over half of the participants had sustained an injury in the past two years, however, this rate was not significantly different between the different recruitment categories. No association was found between eligibility for genetics and recruitment method.

Table 2.3 Respondents' characteristics by recruitment strategy within each category.

Category		Facebook (%)	Other social media (%)	Online media (%)	Running events (%)	*Parkrun* (%)	Referral (%)	AIS (%)	Other (%)
Sex*	Male	38.6	50.9	67.7	51.1	48.9	54.9	50.4	36.4
	Female	61.4	49.1	32.3	48.9	51.1	45.1	49.6	63.6
Age*	18-24	6.8	7.7	4.3	7.1	2.7	5.6	15.0	9.1
	25-34	22	17.6	26.1	25.9	14.4	16.7	25.6	15.2
	35-44	39.1	27.5	30.4	30.6	25.8	34	24.1	33.3
	45-54	22.8	25.3	27.3	25	32.3	29.2	24.1	21.2
	55-64	8.2	17.6	8.7	9.2	17.9	13.9	9	9.1
	≥ 65	1.1	4.3	3.1	2.1	6.7	0.7	2.3	12.1
Weekly running distance*	< 20 km	29.9	32.4	34.8	38.7	43.0	31.9	42.9	43.8
	20-40 km	46.9	42.6	44.1	42.8	42.8	44.4	40.6	34.4
	40 km	23.2	25	21.1	18.5	14.2	23.6	16.5	21.9
Injured in the past two years	Yes	59.1	53.4	54.7	52.6	55.7	62.5	54.9	48.5
	No	40.9	46.6	45.3	47.4	44.3	37.5	45.1	51.5
Eligible for genetic study	Yes	35.2	30.6	33.5	29.9	26.9	30.6	30.8	20.6
	No	64.8	69.4	66.5	70.1	73.1	69.4	69.2	79.4

* - statistically significant difference between categories (p<0.001).

Whilst an online survey platform and described recruitment strategies were utilised in an effort to recruit recreational runners across all states and territories of Australia, the analysis of runners' location by state showed that the most commonly reported residential states were three largest states of Australia: New South Wales (NSW) - 25.2 , Queensland (QLD) – 19.7 and Victoria (VIC) – 17.2 . However, the fourth most reported location was the Australian Capital Territory (ACT) – 11.6 , which is the second smallest state or territory in Australia. The least common states and territories where residents completed the survey were Western Australia (WA) – 6.5 , South Australia (SA) – 3.8 , Tasmania (TAS) – 2.5 , and Northern Territory (NT) – 0.5 . However, 13.1 did not provide their residential state or territory (Figure 2.2).

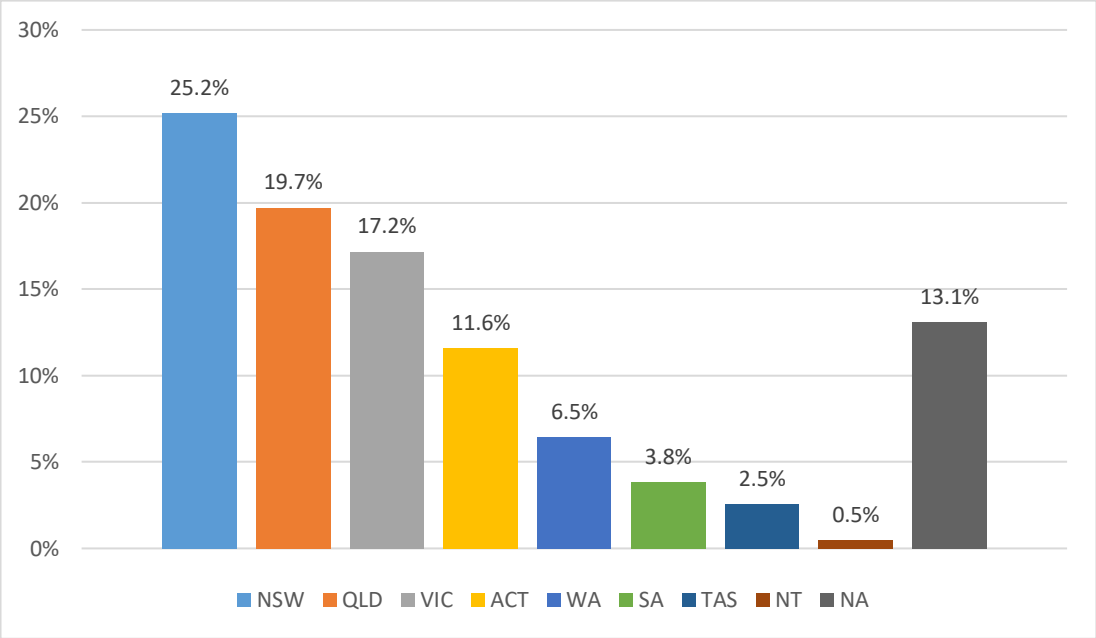


Figure 2.2 Distribution of the respondents by their residential state/territory.
NA – Not Available

In addition, when numbers of respondents from each state were recalculated as per 100,000 people, the ACT had the highest rate of participants per 100,000 territory residents – 142 per 100,000. This number was approximately seven times higher than values in the three largest states of Australia – NSW (n 16), VIC (n 14) and QLD (n 20) (Table 2.4). Although Tasmania had the second-highest rate of respondents per 100,000 residents (n 24), this state was represented only by 125 runners completed the survey. Two remaining states WA

and SA, had lower than other states scores with only 319 and 188 runners representing these states in the study, respectively. The least represented territory was NT with just 23 completed surveys.

Table 2.4 Distribution of respondents by state and rates of participant numbers by 100,000 residents in each state.
Population data sourced from (Population Australia, 2018).

State	Population	Number of runners	Number per 100,000
ACT	401,137	571	142
TAS	519,166	125	24
QLD	4,900,000	974	20
NSW	7,700,000	1244	16
VIC	6,150,000	848	14
WA	2,640,000	319	12
SA	1,710,000	188	11
NT	244,500	23	9

2.4 Results - A profile of health, lifestyle & training habits of 4720 Australian recreational runners

Data from 5,250 respondents who described themselves as recreational runners were collected over the 25 months of recruitment. After duplicate (n 272), nonsense (n 4) and incomplete (n 35) responses were removed, and 4,939 responses remained. As a weekly running distance of greater than 15 km was stated as an inclusion criterion for participation in the survey, data from 219 runners who reported less than this distance were removed, resulting in 4,720 responses included in the analysis. All respondents were 18 years of age or over. The study cohort was 54.1 female and 45.9 male (Table 2.5).

Table 2.5 Physical characteristics of the Australian recreational running cohort.

Characteristics	All runners (*N*=4720)		Male runners (*N*=2165)		Female runners (*N*=2555)	
	Median (range)	**IQR**	**Median (range)**	**IQR**	**Median (range)**	**IQR**
Age (years)	40 (18 – 80)	33 – 47	42 (18 – 80)	34 – 49	39 (18 – 77)	32 – 46
Weight (kg)	68 (40 – 135)	60 – 77	76 (45 – 135)	70 – 83	61 (40 – 110)	56 – 68
Height (cm)	172 (120 – 210)	165 – 179	179 (152 – 210)	175 – 183	166 (120 – 190)	162 – 170
Body Mass Index (BMI) (kg/m^2)	23 (16 – 44.4)	21.3 – 25	23.8 (16 – 40.8)	22.3 – 25.6	22.2 (16.3 – 4.4)	20.6 – 24.2

2.4.1 Running habits

The training characteristics of respondents are described in Table 2.6. The most common weekly running distance was 20–40 km (45.8) among the entire cohort, with very similar rates in male and female runners. However, males were more likely than females to run distances more than 40 km per week (χ_1^2 77.6, p<0.001), whereas females were more likely than males to run 15–20 km per week (χ_1^2 65.5, p<0.001). The most common category of respondents was of those with over ten years of running experience (37.8), with significantly more males than females within this experienced group (χ_1^2 71.3, p<0.001). The majority of respondents stated that they typically ran between two and five sessions per week. It was, however observed that males were significantly more likely to run six or more times per week than females (χ_1^2 33.3, p<0.001). The typical race pace of a male runner was reported as 4–5 min/km, whereas female runners reported 5–6 min/km. The majority of respondents participated in other sports in addition to running. Significantly more female than male runners reported participation in sports other than running (χ_1^2 63.8, p<0.001). Over half of recreational runners reported injuries that occurred while running in the past two years, with significantly higher rates in males than in females (χ_1^2 4.84, p 0.016).

Table 2.6 Training characteristics of the Australian recreational running cohort.

Training characteristics		All runners (*N*=4720)		Male runners (*N*=2165)		Female runners (*N*=2555)	
		n	**%**	***n***	**%**	***n***	**%**
Weekly running distance**	15-20 km	1424	30.2	526	24.3	898	35.1
	20-40 km	2164	45.8	991	45.8	1173	45.9
	40 km	1132	24.0	648	29.9	484	18.9
Running experience**	≤2 years	942	20.0	364	16.8	578	22.6
	3-5 years	1267	26.8	542	25.0	725	28.4
	6-9 years	722	15.1	297	13.7	425	16.6
	10+ years	1783	37.8	958	44.2	825	32.3
	NA	6	0.1	4	0.2	2	0.1
Run sessions per week**	1	15	0.3	10	0.5	5	0.2
	2 or 3	2041	43.2	869	40.1	1172	45.9
	4 or 5	2226	47.5	1027	47.4	1199	46.9
	6+	422	8.9	250	11.5	172	6.7
	NA	16	0.3	9	0.4	7	0.3

NA – Not Available; * - statistically significant difference between males and females ($p<0.05$), **- statistically significant difference between males and females ($p<0.001$).

Table 2.6 Training characteristics of the Australian recreational running cohort (continuation).

Training characteristics		All runners (*N*=4720)		Male runners (*N*=2165)		Female runners (*N*=2555)	
		n	**%**	***n***	**%**	***n***	**%**
Race pace**	<4 min/km	403	8.5	329	15.2	74	2.9
	4-5 min/km	1591	33.7	1022	47.2	569	22.3
	5-6 min/km	1819	38.5	633	29.2	1186	46.4
	6-7 min/km	706	15.0	141	6.5	565	22.1
	7 min km	189	4.0	34	1.6	155	6.1
	NA	12	0.3	6	0.3	6	0.2
Participation in other sports**	Yes	3590	76.1	1530	70.7	2060	80.6
	No	1113	23.6	629	29.1	484	18.9
	NA	17	0.4	6	0.2	11	0.5
Reported injuries occurred while running in past two years*	No	1969	41.7	866	40.0	1103	43.2
	Yes	2751	58.3	1299	60.0	1452	56.8

NA – Not Available; * - statistically significant difference between males and females ($p<0.05$), **- statistically significant difference between males and females ($p<0.001$).

2.4.2 Smoking habits

Smoking was uncommon among surveyed runners, with 0.6 reporting that they were current smokers and a further 25.8 of runners reporting that they had smoked at any time in their life. The reported smoking experience was not significantly associated with sex (24.7 versus 26.7 ; χ_1^2 2.1, p 0.1). The analysis of self-reported rates of alcohol consumption rates has been excluded from this study due to its unreliability, supported by other studies of self-reported alcohol use (Ekholm, Strandberg-Larsen, & Gr nb k, 2011; Proude, Britt, Valenti, & Conigrave, 2006).

2.4.3 Chronic conditions

The survey included questions about 18 lifetime diagnoses of chronic conditions (Table 2.7). The most common reported diagnosis was depression (15.3) with significantly higher reported depression rates among females than males (χ_1^2 55.7, p<0.001). The second most common diagnosis was respiratory conditions (11.7), which was significantly higher in females than males (χ_1^2 9.0, p 0.002). Although anaemia was the third most common diagnosis (10), this was mainly reported by females (χ_1^2 315.8, p<0.001). A lifetime diagnosis of hypertension was reported by 290 runners (6.1). Hypertension was the third most common diagnosis for males, accounting for 7.8 , with a significantly lower reported rate in females – 4.8 (χ_1^2 18.1, p<0.001).

Table 2.7 Lifetime diagnoses of chronic conditions reported by recreational runners.

Chronic conditions	All runners (*N*=4720)		Male runners (*N*=2165)		Female runners (*N*=2555)	
	n	**%**	***n***	**%**	***n***	**%**
Depression**	724	15.3	240	11.1	484	18.9
Respiratory conditions**	554	11.7	221	10.2	333	13.0
Anaemia**	472	10.0	34	1.6	438	17.1
Hypertension**	290	6.1	168	7.8	122	4.8
Skin disease*	277	5.9	111	5.1	166	6.5
Cancer	238	5.0	108	5.0	130	5.1
Insomnia**	193	4.1	54	2.5	139	5.4
Osteoarthritis	184	3.9	74	3.4	110	4.3
Gastrointestinal disease**	186	3.9	56	2.6	130	5.1
Cardiac conditions	179	3.8	94	4.3	85	3.3
Thyroid disease**	180	3.8	30	1.4	150	5.9
Neurological conditions	88	1.9	37	1.7	51	2.0
Diabetes	68	1.4	30	1.4	38	1.5
Rheumatoid arthritis	54	1.1	19	0.9	35	1.4
Osteoporosis*	53	1.1	14	0.6	39	1.5
Chronic renal failure	11	0.2	7	0.3	4	0.2
Cerebral palsy	8	0.2	5	0.2	3	0.1
Cystic fibrosis	7	0.1	5	0.2	2	0.1

* - statistically significant difference between males and females ($p<0.05$), ** - statistically significant difference between males and females ($p<0.001$)

2.4.4 Body mass index and weight loss

Respondents were grouped by their BMI in accordance with the World Health Organisation guidelines (Table 2.8) (World Health Organization, 2000). The majority of runners were in the normal weight category with a BMI between 18.5 kg/m^2 and 25 kg/m^2 (72.9). Of the remainder, 2.6 of runners were underweight (16 to 18.5 kg/m^2), 21.8 were overweight (25 to <30 kg/m^2) and 2.7% were obese (≥30 kg/m^2). There were no participants in the cohort that were classified as severely underweight (<16 kg/m^2). When levels of obesity were categorised, 127 runners were divided into 3 subgroups: moderate level of obesity (*n* 111), severe (*n* 14) and very severe (*n* 2), accounting for approximately 2.35 , 0.30 and 0.04 , respectively, of the entire sample. Due to these low numbers in severe and very severe categories, it was decided to keep a general 'obese' group for analysis. Significantly more women than men were in the underweight and normal weight categories (χ_1^2 50.6, p<0.001; χ_1^2 66.8, p<0.001, respectively), whereas significantly more men than women were overweight (χ_1^2 136.3, p<0.001). Nevertheless, the proportions of obese male and female runners were almost equal. The BMI distribution data of surveyed runners were compared to BMI data in the Australian population collected in the 2014-15 National Health Survey published by the Australian Bureau of Statistics (Figure 2.3) (Australian Bureau of Statistics, 2016a).

Table 2.8 Frequencies of the body mass index categories.

Body Mass Index	All runners (*N*=4720)		Male runners (*N*=2165)		Female runners (*N*=2555)	
	n	**%**	***n***	**%**	***n***	**%**
Underweight (<18.5 kg/m^2)*	121	2.6	17	0.8	104	4.1
Normal (18.5 to <25 kg/m^2)*	3443	72.9	1455	67.2	1988	77.8
Overweight (25 to <30 kg/m^2)*	1029	21.8	637	29.4	392	15.3
Obese (≥30 kg/m^2)	127	2.7	56	2.6	71	2.8

* - statistically significant difference between males and females (p<0.001)

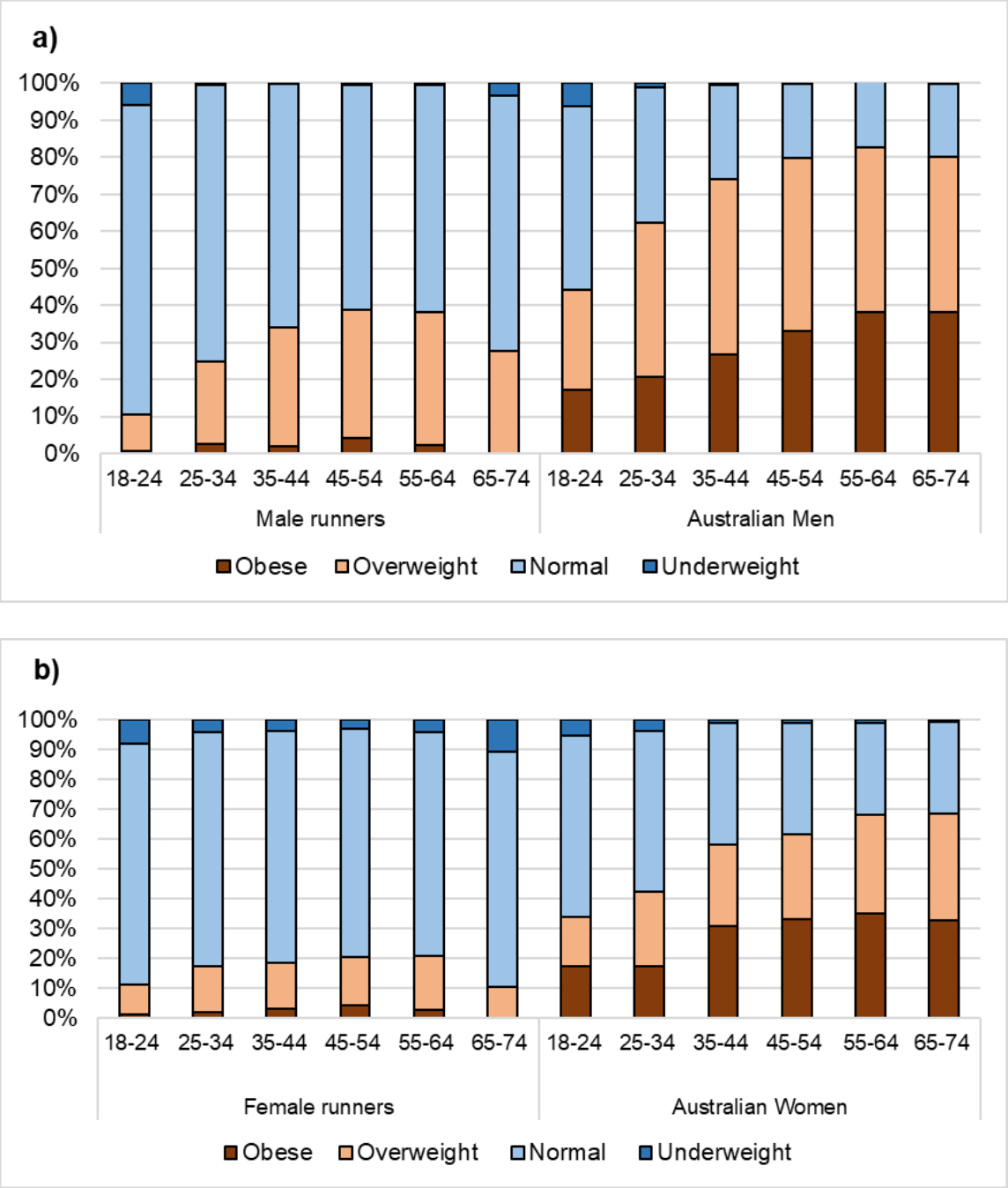


Figure 2.3 Comparison of BMI group percentages between Australian surveyed a) male and b) female runners and Australian population surveyed by the Australian Bureau of Statistics of different age groups.

Recreational runners were asked whether they had gained or lost a clinically significant amount of weight (≥5kg) in the past two years. Clinically significant weight loss over the last two years was reported by 27 of all respondents. Multiple logistic regression analysis (Table 2.9) showed that clinically significant weight loss was more likely to be reported by younger runners, and overweight and obese runners. Runners with two or fewer years of running experience were three times more likely to report clinically significant weight loss in the past two years than runners with over ten years of running experience. However, sex, participation in other sports, and history of injuries in the past two years did not have a statistically significant association with clinically significant weight loss. Interestingly, smoking experience (smoking at any time in life) was associated with reported clinically significant weight loss. The logistic regression results indicate that commencing a running program may lead to a clinically significant weight loss irrespective of sex, participation in other sports and injury in the previous two years.

Table 2.9 Multiple logistic regression analysis results with adjusted odds ratio (OR) estimates for the effects of runner characteristics on clinically significant weight loss (≥5kg).

Variable	OR[1]	95% CI	*p*-value
Sex			
Female[2]	1.00		
Male	1.04	0.90, 1.19	0.62
Age group			
55 years[2]	1.00		
35-55 years	1.91	1.43, 2.54	<0.001
< 35 years	2.20	1.62, 2.96	<0.001
BMI group			
Normal[2]	1.00		
Underweight	0.69	0.42, 1.13	0.14
Overweight	1.96	1.68, 2.29	<0.001
Obese	2.51	1.76, 3.58	<0.001
Running experience			
≥ 10 years[2]	1.00		
6-9 years	1.22	0.98, 1.52	0.07
3-5 years	1.63	1.36, 1.94	<0.001
≤ 2 years	3.15	2.63, 3.78	<0.001
Participation in other sports			
No[2]	1.00		
Yes	0.99	0.85, 1.16	0.92
Injury occurrence			
No[2]	1.00		
Yes	1.02	0.89, 1.69	0.77
Smoking history			
No[2]	1.00		
Yes	1.34	1.15, 1.56	<0.001

CI – Confidence Interval, [1] - Adjusted for the other variables in the table; [2] - Reference category

2.5 Discussion

2.5.1 Discussion of the recruitment strategies and participants' residential distribution

The development of secure online platforms for data collection, the ubiquitous presence of the Internet and the multiplicity of access devices allow researchers to collect health data for multiple research purposes. The online survey utilised for this project was completed by over 5,000 Australian recreational runners who were half of the initially targeted number of 10,000 runners. The number of adult runners in Australia in 2015-2016 was estimated at over 2.8 million (Australian Sports Commission, 2016), hence there was no shortage of potential participants. The analyses of the recruitment strategies employed, their efficiency and their impact on population representativeness identified some associations between participants' characteristics and recruitment strategies. Thus, Facebook and online media had an effect on participants' sex distribution. Despite the absence of sex-related differences in social media or Facebook usage or behaviour (Sensis, 2016), Facebook was more likely to recruit women to the survey than men. The female recruitment rate through Facebook (61.39) aligned with previous studies, which also used Facebook (60 on average) as summarised in a recent systematic review (Thornton et al., 2016). The opposite was true for online media, however, the leading online media channel used was a specific blog with a predominantly male readership.

Although Facebook continues to be the most popular social media platform in Australia for all ages (Sensis, 2016), this platform preferentially recruited participants in the 35-44 age bracket. This bias may be a result of the advertisement campaign being targeted exclusively to runners aged 30-50 during the first 3 months of the 9-month long campaign, although it was later extended to ages 18-50 for a further 6 months. The runners recruited from *parkrun*, a network of free, weekly, timed 5 km runs in outdoor spaces, were predominantly in the 55-64 age category. This age group was also mainly recruited by other social media, particularly our email campaign, which spread information through media such as members' newsletters of *parkrun* and running clubs. Interestingly, the youngest age group, 18-24 years, were more likely to be recruited directly through the AIS. Possibly, since the AIS provides programs and facilities for developing and elite athletes, younger survey respondents who were initially attracted to the AIS website were young athletes. Respondents who reported running less than 20 km per week were over-represented and those who run over 40 km per week were under-represented in the *parkrun* group, which aligns with the observation that *parkrun* attracts people with the lower running ability (Stevinson & Hickson, 2013).

Additional analysis of the residential location of the respondents demonstrated recruitment bias towards Eastern shore states and territories - QLD, NSW, ACT and VIC. The highest rate of respondents per 100,000 people was identified in the ACT. This highest rate could be explained by the fact that the research group responsible for the recruitment was based in the ACT and therefore had access to more recruitment opportunities at the local running events. In addition, the AIS is a renowned institution in the ACT, which could also help to draw more interest to the project from the local physically active community. Running and media events in the remaining three states were attended by the research group for promotion purposes multiple times. This attendance may have boosted recruitment in these particular states, whereas more remote states and territories were reached only through online social media resources and emails.

Overall, the data suggest that the combined use of traditional, online and physical recruitment strategies resulted in a diverse sample. However, the recruitment bias towards older age among physically active people suggests that people aged under 35 are particularly hard to reach and may require the employment of more specific recruitment strategies. Additionally, offline recruitment contributes to the bias of the geographical distribution of the recruited participants. This suggests that if research requires a particular distribution of participants across the country, online and offline strategies should be implemented evenly across all states and territories.

2.5.2 Discussion of physical characteristics, training and lifestyle habits of 4,720 Australian recreational runners

This study described one of the largest cohorts of Australian recreational runners, analysing the medical and lifestyle characteristics of the participants and sex differences in training habits. In this study, we demonstrated that a large proportion of recreational runners avoided the majority of modifiable risk factors that contribute to the burden of disease. In the Australian population, the five strongest contributors to the burden of disease in 2011 were tobacco use (9), high body mass (5.5), alcohol use (5), physical inactivity (5) and hypertension (5) (Australian Institute of Health and Welfare, 2016a). Data from the 'Australian Nutrition and Physical Activity Survey' showed that both sufficient physical activity level and reduced sitting time were important factors for the prevention of cardiovascular disease and metabolic syndrome (Engelen et al., 2016). This study demonstrated that Australian recreational runners typically have a BMI in the normal range, are meeting physical activity guidelines through recreational running and participation in other sports and have low levels of smoking. This cohort runs on average 20–40 km in greater than two sessions per week, and 76.1 of respondents play additional sport, indicating that recreational runners are likely to be meeting the recommended 'WHO Physical Activity Guidelines' (World Health Organisation, 2010). Considering that 80 of surveyed recreational runners have been running for at least three years, we can speculate that they have managed to sustain a habit of regular physical activity at the recommended level for at least three years.

Australian recreational runners self-reported a lower BMI than the general population. Additionally, a weight loss of greater than five kilograms in the past two years was reported by approximately 40 of runners with less than two years of experience. Physical activity is a critical component in the multidisciplinary approach of effective weight loss programs, and is especially important when preventing continued weight gain or maintaining lower weight (S derlund, Fischer, & Johansson, 2009). Indeed, endurance running has been shown to be beneficial to physically inactive adults leading to body mass and body fat reduction, with a systematic review concluding that one year of running training was effective in reducing body mass by 3.3 kg (Hespanhol Junior, Pillay, van Mechelen, & Verhagen, 2015). Several systematic reviews have shown that aerobic exercise, such as running, significantly contributes to weight loss, with strong evidence that this type of activity is effective in reducing visceral fat (Ismail, Keating, Baker, & Johnson, 2012; Thorogood et al., 2011).

Additionally, there is a dose-response relationship between aerobic exercise and visceral fat reduction in obese participants, indicating that an activity such as recreational running could be effective in improving health via a reduction of visceral fat (Ohkawara, Tanaka, Miyachi, Ishikawa-Takata, & Tabata, 2007). A systematic review indicated that risks of all-cause mortality and cardiovascular mortality were lower in people with high BMI and good aerobic fitness than in people with normal BMI and poor fitness. However, aerobically fit people with high BMI were still at a greater risk of type 2 diabetes mellitus and cardiovascular disease (Fogelholm, 2010). In an Australian population, it has been shown that walking is the most common type of physical activity recommended to patients by their doctor (Porter, Eccleston, & Vilshanskaya, 2002; Robertson, Jepson, Shepherd, & McInnes, 2011). Our results, taken together with previous findings, indicate that recreational running could be promoted by general practitioners as an effective mechanism for building aerobic fitness and maintaining healthy body weight.

Only one-quarter of recreational runners surveyed reported smoking at any time during their life, and 0.6 were current smokers. These rates were substantially lower than those reported by the Australian Bureau of Statistics, which showed that 14.5 of adult Australians were daily smokers, 1.5 smoked less often than daily, and about one third (31.4) were ex-smokers (Australian Bureau of Statistics, 2016a). A systematic review of co-occurrence of smoking and physical activity showed negative association in 20 studies on adults in several European countries, Japan and Australia and 13 studies with nonsignificant, mixed or positive association, indicating possible complex relationships between smoking and physical activity due to race, income level and other factors (Kaczynski, Manske, Mannell, & Grewal, 2008). Hence, the very small proportion of runners currently smoking could reflect their overall healthy lifestyle as well as the positive effects of individual exercise bouts in reducing cravings for smoking.

Depression was the most common life-time diagnosis reported by recreational runners (15.3). Affective disorders which comprise all levels and severity of depressive disorders and bipolar disorder, accounted for 15 of life-time prevalence in the Australian population (Slade, Teesson, & Burgess, 2009). This study population contained a slightly higher proportion of females (54.1) compared to the general Australian population (50.4) (Australian Institute of Health and Welfare, 2018). As female recreational runners were more likely to report depression, this finding was similar to results of other studies (Kruijshaar et

al., 2005; Nolen-Hoeksema, 2001). Overall, the depression rate in the Australian recreational running population is largely the same as in the general population.

Hypertension is a significant risk factor for chronic diseases including stroke, coronary heart disease, heart failure and chronic kidney disease and is identified as the leading global risk factor for mortality (World Health Organization, 2009). Based on measured data from the Australian Institute of Health and Welfare, 32 of Australians aged 18 and over have hypertension (Australian Institute of Health and Welfare, 2017). However, only 6.1 of Australian recreational runners surveyed self-reported that they had been diagnosed with hypertension. The reduced levels suggest that recreational running is associated with lower rates of hypertension as a risk factor for burden of disease. These findings are supported by the previous reports of approximately 50 reduction in the risk of hypertension in long-distance runners (Williams, 1997), and a meta-analysis of 72 studies of effect of aerobic endurance exercise (running, cycling, swimming) on blood pressure, demonstrated that this type of physical activity reduces blood pressure, and the reduction was more pronounced in hypertensives than non-hypertensives (Cornelissen & Fagard, 2005).

Running, as a form of physical activity, has consistently been shown to provide a range of health benefits, including reducing the overall risk of cardiovascular disease and all-cause mortality (Chakravarty, Hubert, Lingala, & Fries, 2008; Lee et al., 2014; Oja et al., 2016; Schnohr, O'Keefe, Marott, Lange, & Jensen, 2015). More importantly, the clustering of various healthy behaviours has been shown to be inversely related to the risk of all-cause mortality, with four or more healthy behaviours reducing mortality risk by 66 (Loef & Walach, 2012). Here we demonstrated that a large proportion of Australian recreational runners displayed healthy behaviours including meeting physical activity guidelines, avoidance of overweight or obesity and reduced smoking.

We suggest that recreational running could be promoted as a low-cost option for adhering to physical activity guidelines. Marketing of recreational running through mass participation events, for example, *parkrun* has been considered as a public health intervention (Stevinson & Hickson, 2013; Stevinson, Wiltshire, & Hickson, 2015). However, there is a risk of sustaining an injury during participation in recreational running, with 58.3 of participants in the 'AIS Running Injury Study' reporting a running injury over the preceding two years. A recent study of 1,145 *parkrun* participants in the United Kingdom reported a 49.8 injury rate over 12 months (Linton & Valentin, 2018). This potential injury risk must be considered when advising participation in recreational running. Additionally, the commencement of a

running program for individuals with musculoskeletal injuries of the lower body should be supervised by a qualified medical practitioner. The current study demonstrated that there are differences in male and female training characteristics to consider when aiming to encourage people to begin a running program. We show here that female runners were more likely to report shorter weekly distances while running a similar number of sessions as male runners. This study did not investigate the motivations for participation in recreational running, however several studies have demonstrated that the motivations of males and females, in relation to participation in physical activity, differ in a number of ways (Lauderdale, Yli-Piipari, Irwin, & Layne, 2015; Louw, Van Biljon, & Mugandani, 2012; Stults-Kolehmainen, Ciccolo, Bartholomew, Seifert, & Portman, 2013). An Australian study demonstrated that, while both males and females are motivated by general health and maintenance of fitness, women often cite weight loss/appearance and mental health as motivating factors for increasing their physical activity levels while men participate for social reasons and enjoyment (Australian Sports Commission, 2016). Both motivational factors and running habits should be considered when marketing recreational running for health benefits or encouraging participation.

While the self-report nature of data collection could introduce bias and error, the survey tool has been shown to be reliable and questions did not require respondents to recall long-term details of running habits or injuries (Domaschenz et al., 2015). The term of recall for injuries and running habits was limited to the two years preceding survey response, as it has been shown that retrospective data beyond this point is not reliable (Gabbe, Finch, Bennell, & Wajswelner, 2003; Kolt & Kirkby, 1999). A further study limitation may have been sampling bias with a higher proportion of female runners and middle-aged runners in the studied cohort in comparison with demographic data of physically active Australian adults (Australian Bureau of Statistics, 2015). Lastly, the absence of data from non-runners or those who may be interested in taking up recreational running precludes extrapolation of findings to non-runners.

Recreational running is associated with benefits across a range of measurable health outcomes. A high proportion of the Australian recreational runners who participated in this study had a body mass index within the healthy weight range, seemed to be meeting the WHO Physical Activity Guidelines each week for many years and were non-smokers. Additionally, our results indicate that taking up running is associated with weight loss and weight remains stable if individuals persist with running. Male and female runners reported

different running characteristics, and these should be considered when promoting recreational running or encouraging participation.

3. Chapter Three – Training-related factors associated with running-related injuries, and specifically Achilles tendon injuries and bone stress injuries

Addendum

Contributions to Chapter 3:

Mariia Kozlovskaia:

- Recruitment of participants
- Data collection
- Database management
- Data quality control
- Statistical data analyses
- Author of the chapter

Nicole Vlahovich:

- Online survey design and development
- Ethics applications
- Recruitment of participants
- Database management
- Editing of the chapter

Evelyne Rathbone:

- Assistance with statistical analyses

3.1 Introduction

Running is one of the most popular recreational physical activities in the Australian population (Australian Sports Commission, 2016). Running is a low cost and an easily implemented activity with proven health benefits involving improvements in aerobic fitness and cardiovascular function (Oja et al., 2015). However, running is a sport with relatively high rates of injuries, varying between 19 and 79 (van Gent et al., 2007). Running-related injuries may be of acute or overuse nature, and the latter comprises 80 , mainly affecting Achilles tendon, knee and foot (Walther et al., 2005). According to a systematic review of running-related injuries, three most common injuries were: medial tibial stress syndrome (13.6—20), Achilles tendinopathy (9.1—10.9) and plantar fasciitis (4.5—10) (Dias Lopes et al., 2012). The development of running-related injuries is affected by a number of modifiable and non-modifiable factors, of which interrelationship is highly complex.

Among physical characteristics, factors such as age, sex, BMI were extensively studied on multiple running cohorts. However, all these factors showed inconsistent results across many studies, confirming the complexity of the risk of injuries, and challenges of research design to study risks factors (van der Worp et al., 2015; van Gent et al., 2007; Van Mechelen, 1992). Training habits of runners are modifiable factors, which may either increase or decrease the risk of running-related injuries. Thus, a weekly running distance of more than 64 km was reported a risk factor of injuries in two studies (Macera et al., 1989; Walter et al., 1989). Additionally, high-frequency running sessions was identified as a risk factor in one study (Walter et al., 1989), whereas a later study showed that one running session per week may increase risk of injuries in female runners (Taunton et al., 2003). Stretching was another risk factor of running injuries investigated in several studies but showed ambiguous results (Hreljac, 2005; Shrier, 1999). Wearing orthotics and a certain type of running shoes is important for the running biomechanics, and their correct choice may assist in injury prevention (McKenzie et al., 1985; Nigg et al., 1999). One of the main identified factors was a previous injury occurred in the past 12 months, however, this association could be explained by incomplete recovery from the original injury (Saragiotto et al., 2014; Walter et al., 1989).

Risk factors of different types of running-relates injuries are diverse and require more specific research. This study was focused on the two most common running-related injuries: an Achilles tendon injury and a bone stress injury. According to the literature review in

Chapter 1, male sex, older age and training overload were the main risk factors of Achilles tendon injuries, however, these factors may be interdependent, as males would typically report higher training load than females (Dias Lopes et al., 2012; Hirschm ller et al., 2012; Magnan et al., 2014; Taunton et al., 2002). The main previously identified risk factors of bone stress injuries were female sex, older age, increased BMI, training regimen, running surfaces and wearing orthotics (Beck et al., 2015; Bennell et al., 1999; Newman et al., 2013; Warden et al., 2014). However, these factors also interact and may cumulatively contribute to the development of bone stress injuries. Overall, the multifactorial nature of running-related injuries makes their research challenging and requires careful and comprehensive data analyses to identify important risk factors.

This part of the study aimed to identify the most common reported running-related injuries and their frequencies, and investigate risk factors of running-related injuries, particularly Achilles tendon injuries and bone stress injuries, reported by Australian recreational runners from the 'AIS Running Injury Study' cohort.

3.2 Methods

3.2.1 Selection criteria for running-related injury study

Recruitment efforts described in Chapter 2 resulted in a dataset of 4,720 unique responses from Australian recreational runners. All included participants were at least 18 years of age and reported running a minimum of 15 km per week. In the survey, runners were asked whether they acquired any injuries which affected their running routine in two years prior to the survey completion. Injured runners could describe a maximum of four injuries that occurred in the past two years. The questionnaire comprised questions specific to Achilles tendon injuries and bone stress injuries in the lower leg. All other types of injuries were reported via open-ended questions and then were classified by either a formal injury diagnosis, for example 'plantar fasciitis' or 'iliotibial band syndrome' (if a participant specified whether this injury was diagnosed by a health professional or clearly detailed the diagnosis), or their anatomic location, if provided description of the injury was not precise. Additional questions about how the reported injuries occurred allowed the exclusion of not running-related injuries and the analysis of only injuries which occurred while running. Therefore, only uninjured runners and runners who reported running-related injuries were included in the Running Injury (RI) cohort and the subsequent statistical analysis. The survey contained questions specifically about Achilles tendon injuries and bone stress injuries, as they were the injuries of interest in the project. Hence, besides a larger cohort of uninjured runners and runners with any type of running-related injury, two additional cohorts were formed: AT injury cohort, which comprised Achilles tendon injury cases and all uninjured runners, and BS injury cohort, which comprised bone stress injury cases and all uninjured runners.

3.2.2 Statistical data analyses

Categorical variables which describe physical and training characteristics of recreational runners were statistically compared using chi-squared χ^2 test in each of three created cohorts (RI, AT, BS). The results were summarised and described in the tables in Section 3.3. A multiple logistic regression method was selected to identify which factors are independently predictive of the outcome – an injury, in the presence of the other factors. Firstly, simple logistic regression using an injury as the outcome was conducted to identify univariate predictors with a significance level of $p \leq 0.1$ to include in the subsequent multiple logistic regression. This significance threshold of $p \leq 0.1$ was selected in order to identify variables that were not independently statistically significant but at the presence of other

variables in the multiple logistic regression model. Thus, these variables were included in the logistic regression analysis and underwent a number of iterations. During each iteration, the least significant variable was excluded from the model until the final logistic regression model included only statistically significant variables. Hosmer-Lemeshow test was used to assess the quality of the models with p 0.05, indicating an acceptable model fit. Logistics regression models with significant variables were summarised in the tables for each of three investigated cohorts.

3.3 Results

3.3.1 Frequencies of reported running-related injuries

After the selection of the injuries, which occurred while running, 436 runners with injuries irrelevant to running were excluded, and a Running Injury (RI) cohort (N 4,284) was formed and comprised 1,969 (46) uninjured runners and 2,315 (54) runners, who reported various running-related injuries occurred in the past two years. The RI cohort was used to identify and classify running-related injuries and report frequencies of the injuries in the whole cohort and male and female runners separately. Injured runners were able to report between one and four running-related injuries that may belong to different groups, as well as recurrent injuries, and occurred in the past two years. Among 2,315 injured runners, 1,360 (58.7) reported only one injury, 749 (32.4) reported two injuries, 162 (7) runners reported three injuries, and 44 (1.9) runners reported four injuries, resulting in 3,520 reported injuries in total. Among these 3,250 injuries, 279 (7.9) were either traumatic or unidentified running injuries; that is, the injury was reported, but no details were provided about the type or location of the injury. The remaining 3,241 injuries were classified depending on the affected anatomic part of the leg: an Achilles tendon injury, a bone stress injury (below knee), iliotibial band (ITB) syndrome, plantar fasciitis, hip and glute injuries, an upper leg injury (including quad tear and femur stress injury), a hamstring injury, a knee injury (including patella injury, meniscus injury, anterior cruciate ligament (ACL) tear), a lower leg injury (posterior tendon injury and unspecified tendinopathies, a calf injury (including muscle strain), an ankle injury (including ligament sprain), foot and toe (including unspecified tendinopathies, bursitis). Two most common of all 3,520 reported injuries were an Achilles tendon injury and a bone stress injury (16 and 15.7 , respectively). A calf injury, a knee injury and ITB syndrome were 12.3 , 10 and 7.3 of all reported injuries,

respectively. Plantar fasciitis, a hamstring injury, an ankle injury, hip and gluteal injuries were of similar frequencies just above 5 (5.5 , 5.3 , 5.2 and 5.1 , respectively). Foot and toe injuries comprised only 4.2 . Upper leg injuries and lower leg injuries were 1.6 and 3.9 , respectively (Figure 3.1). Overall, 62.8 of all reported injuries affected the lower leg, 17.3 affected knee area (knee injury and ITB syndrome), 12 of injuries affected upper leg, including hip and glute, and 7.9 of injuries were unspecified.

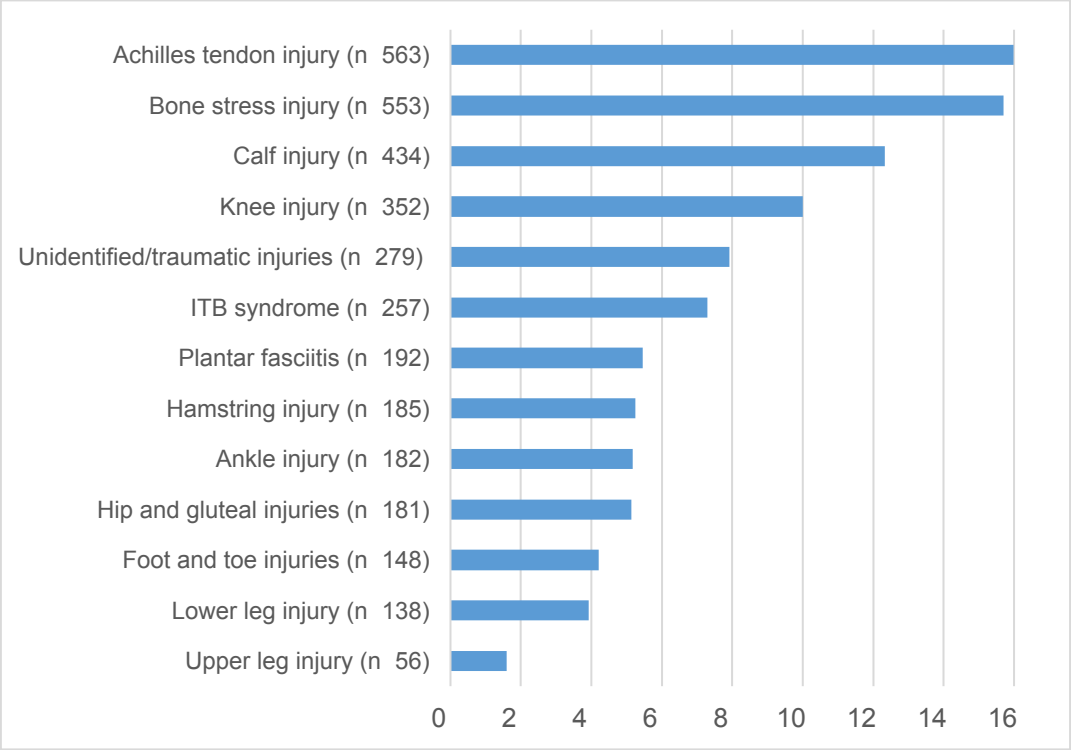


Figure 3.1 Frequencies of reported running-related injuries.
*ITB syndrome – iliotibial band syndrome.

Runners also reported the number of weeks they interrupted running practice after each injury. However, these data were provided only for 3,484 injuries out of 3,520, resulting in 1 (n 36) missing data. Only 7.4 of injuries prevented runners from training for one week. Over half of the reported injuries led to two to four weeks off running (56.8), 11.6 of the injuries interrupted running practice for five or six weeks, whereas the remaining 23.2 of the injuries led to a period of longer than six weeks to return to running (Table 3.1).

Table 3.1 Frequencies of reported weeks off running due to an injury.

Number of weeks off running	Injuries (*N*=3520)	
	n	**%**
1	261	7.4
2	812	23.1
3	604	17.2
4	582	16.5
5	142	4.0
6	265	7.5
6+	818	23.2
NA	36	1.0

NA – Not Available

3.3.2 Physical characteristics and training factors associated with running-related injuries

In order to investigate the risk factors associated with running-related injuries, physical characteristics and training habits of 2,315 injured and 1,969 uninjured runners in the RI cohort were statistically compared (Table 3.2 and Table 3.3, respectively). Injured and uninjured subgroups statistically differed by the proportions of male and female runners (χ^2_1 6.59, p 0.01). All runners were categorised into three age groups: 18-34 years, 35-50 years and over 50 years of age. Injured and uninjured subgroups were significantly different by age group distribution (χ^2_2 7.52, p 0.02), with fewer runners aged over 50 years in the injured group (16.5) versus the uninjured group (19.7). Comparison of injured and uninjured runners by BMI groups (normal, underweight, overweight and obese) showed no statistical difference (χ^2_3 4.83, p 0.19). When injured and uninjured runners were compared by their training habits, there was no significant difference between these subgroups by weekly running distance (χ^2_2 0.57, p=0.75) and reported race pace (χ^2_4 8.89, p 0.06). However, the injured subgroup statistically differed from the uninjured subgroup by years of running experience (χ^2_4 43.02, p<0.001), number of running sessions per week (χ^2_2 11.71, p=0.003) and preferred running terrain (χ^2_6 19.28, p 0.004). In addition, injured runners were more likely to report incorporation of stretching in their training routine than uninjured runners (65.7% versus 58.6%, χ^2_1 23.09, p<0.001), as well as wearing orthotics while running (24.4% versus 13.8%, χ^2_1 77.18, p<0.001), and participation in sports other than running (76.4% versus 73.5%, χ^2_1 4.46, p 0.04). In summary, frequencies of eight variables statistically differed in injured and uninjured subgroups.

Table 3.2 Physical characteristics of uninjured and injured runners.

Physical characteristics		Uninjured runners (*N*=1969)		Injured runners (*N*=2315)	
		n	**%**	***n***	**%**
Sex*	Male	866	44	1109	47.9
	Female	1103	56	1206	52.1
Age groups*	18-34 years	600	30.5	722	31.2
	35-50 years	982	49.9	1212	52.4
	50 years	387	19.7	381	16.5
Body mass index (BMI) groups	Normal	1465	74.4	1701	73.5
	Underweight	57	2.9	53	2.3
	Overweight	405	20.6	492	21.3
	Obese	42	2.1	69	3

* - statistically significant difference between uninjured and injured subgroups (p<0.05).

Table 3.3 Training characteristics of uninjured and injured runners.

Training characteristics		Uninjured runners (*N*=1969)		Injured runners (*N*=2315)	
		n	**%**	***n***	**%**
Weekly running distance	15-20 km	559	28.4	678	29.3
	20-40 km	925	47	1063	45.9
	40+ km	485	24.6	574	24.8
Race pace	<4 min/km	163	8.3	213	9.2
	4-5 min/km	640	32.5	830	35.9
	5-6 min/km	781	39.7	872	37.7
	6-7 min/km	306	15.5	312	13.5
	7 min km	74	3.8	82	3.5
	NA	5	0.3	6	0.3
Running experience**	≤ 1 year	252	12.8	163	7.1
	2 years	197	10.0	242	10.5
	3-5 years	505	25.6	652	28.2
	6-9 years	280	14.2	384	16.6
	10+ years	732	37.2	871	37.6
	NA	3	0.2	3	0.1
Run sessions per week*	1 to 3	852	43.3	999	43.1
	4 or 5	897	45.6	1128	48.7
	6+	211	10.7	182	7.9
	NA	9	0.5	6	0.3
Running terrain*	Bitumen	883	44.8	1060	45.8
	Cement	628	31.9	680	29.4
	Hard dirt/gravel	342	17.4	442	19.1
	Grass	42	2.1	64	2.8
	Treadmill	53	2.7	50	2.2
	Synthetic	6	0.3	16	0.7
	Sand	14	0.7	3	0.1
	NA	1	0.1	0	0
Stretching before/after running**	Yes	1153	58.6	1521	65.7
	No	811	41.2	789	34.1
	NA	5	0.3	5	0.2
Wearing orthotics**	Yes	271	13.8	566	24.4
	No	1689	85.8	1740	75.2
	NA	9	0.5	9	0.4
Participation in other sports*	Yes	1448	73.5	1768	76.4
	No	513	26.1	539	23.3
	NA	8	0.4	8	0.3

NA – Not Available; * - statistically significant difference between uninjured and injured runners (p<0.05), ** - statistically significant difference between uninjured and injured runners (p<0.001).

3.3.3 Logistic regression model of the factors associated with running-related injuries

In order to investigate which of the factors are independently associated with running-related injuries in the presence of other factors, a logistic regression model was built. Firstly, simple logistic regression analysis identified variables of $p \leq 0.1$: sex (p 0.01), age group (p 0.02), years of running experience ($p<0.001$), number of running sessions per week (p 0.003), running terrain (p 0.009), stretching ($p<0.001$), wearing orthotics ($p<0.001$), and participation in other than running sports (p 0.035). After exclusion of non-significant variables from the logistic regression model, six variables remained (Table 3.4). In this model, male sex was associated with 29 increased chance of reporting running-related injuries, runners aged over 50 years were 26 less likely to report running-related injuries than runners aged under 35 years. Runners with one year or less of running experience were also 45 less likely to report injuries when compared with runners with ten and more years of running experience. However, this could be explained by the survey design, which requested injury history of the past two years, which was longer than running experience of these runners. Runners, who reported running one to three sessions and four to five sessions per week, were more likely to report injuries than runners who reported running at least six times per week (34 and 43 , respectively). Additionally, stretching in relation to a running session was associated with a 38 higher chance of reporting running-related injuries. Runners who reported wearing orthotics were also 29 more likely to report running-related injuries.

Table 3.4 Logistic regression model of factors associated with running-related injuries.

Variable	OR (95% CI)	*p*-value
Sex (0=Female, 1=Male)	1.29 (1.13 to 1.46)	<0.001
Age group		
<35 years (reference category)	1.00	
35-50 years	1.00 (0.87 to 1.16)	0.97
50 years	0.74 (0.62 to 0.9)	0.002
Years of running experience		
≤1 year	0.55 (0.44 to 0.69)	<0.001
2 years	1.03 (0.83 to 1.29)	0.77
3-5 years	1.07 (0.92 to 1.26)	0.39
6-9 years	1.15 (0.95 to 1.39)	0.14
≥10 years (reference category)	1.00	
Number of running sessions per week		
1-3 sessions	1.34 (1.07 to 1.68)	0.01
4-5 sessions	1.43 (1.14 to 1.79)	0.002
≥6 sessions (reference category)	1.00	
Stretching in relation to a running session (0=No, 1=Yes)	1.38 (1.13 to 1.46)	<0.001
Wearing orthotics (0=No, 1=Yes)	1.29 (1.13 to 1.46)	<0.001

OR – odds ratio; CI – confidence interval; Hosmer-Lemeshow test χ^2 7.51, p 0.48.

3.3.4 Factors associated with Achilles tendon injuries

One of the most commonly reported injuries was an Achilles tendon injury reported by 508 recreational runners. The AT injury cohort comprised these 508 runners who reported at least one Achilles tendon injury that had occurred in the past two years, and 1,969 uninjured runners. Statistical comparisons of uninjured runners and runners with Achilles tendon injuries identified several significant factors among physical characteristics and training habits (Table 3.5; Table 3.6). Almost two-thirds of runners with Achilles tendon injuries were males. Thus, uninjured runners significantly differed from runners with Achilles tendon injuries by the proportions of male and female runners (χ^2_2 59.69, p<0.001). Uninjured runners also significantly differed from the injured runners by the age group distribution (χ^2_2 19.01, p<0.001). However, there was no statistical difference in the distribution of BMI categories between uninjured and injured runners (χ^2_3 4.58, p 0.21). Injured runners did not differ from uninjured runners by two training characteristics: reported weekly running distance (χ^2_2 3.65, p 0.16) and a number of running sessions per week (χ^2_2 2.66, p 0.26). However, injured runners were significantly different from uninjured runners by frequencies

of runners with different running experience (χ^2_4 38.6, p<0.001), race pace (χ^2_4 26.03, p<0.001) and preferred running terrain (χ^2_6 12.64, p 0.049). Compared subgroups were also statistically different by incorporation of stretching into running practice and wearing orthotics with higher rates of positive responses in the injured subgroup (χ^2_1 7.89, p 0.005, χ^2_1 26.42, p<0.001). However, the rate of participation in other sports was similar for both subgroups (χ^2_1 0.17, p 0.68).

Table 3.5 Physical characteristics of uninjured runners and runners with Achilles tendon injuries.

Physical characteristics		**Uninjured runners (*N*=1969)**		**AT Injured runners (*N*=508)**	
		n	**%**	***n***	**%**
Sex**	Male	866	44	321	63.2
	Female	1103	56	187	36.8
Age groups**	18-34 years	600	30.5	105	20.7
	35-50 years	982	49.9	291	57.3
	50 years	387	19.7	112	22
Body mass index (BMI) groups	Normal	1465	74.4	361	71.1
	Underweight	57	2.9	11	2.2
	Overweight	405	20.6	121	23.8
	Obese	42	2.1	15	3

AT – Achilles tendon; ** - statistically significant difference between uninjured and injured runners (p<0.001).

Table 3.6 Training characteristics of uninjured runners and runners reported Achilles tendon injuries.

Training characteristics		Uninjured runners (*N*=1969)		AT Injured runners (*N*=508)	
		n	**%**	***n***	**%**
Weekly running distance	15-20 km	559	28.4	124	24.4
	20-40 km	925	47	245	48.2
	40+ km	485	24.6	139	27.4
Run sessions per week	1 or 3 sessions	852	43.3	215	42.3
	4 or 5 sessions	897	45.6	248	48.8
	6+ sessions	211	10.7	44	8.7
	NA	9	0.5	9	0.5
Running experience**	≤ 1 year	252	12.8	26	5.1
	2 years	197	10	36	7.1
	3-5 years	505	25.6	120	23.6
	6-9 years	280	14.2	82	16.1
	10+ years	732	37.2	244	48
	NA	3	0.2	0	0
Race pace**	<4 min/km	163	8.3	66	13
	4-5 min/km	640	32.5	202	39.8
	5-6 min/km	781	39.7	166	32.7
	6-7 min/km	306	15.5	60	11.8
	7 min km	74	3.8	14	2.8
	NA	5	0.3	0	0
Running terrain*	Bitumen	883	44.8	234	46.1
	Cement	628	31.9	139	27.4
	Hard dirt/gravel	342	17.4	108	21.3
	Grass	42	2.1	17	3.3
	Treadmill	53	2.7	8	1.6
	Synthetic	6	0.3	1	0.2
	Sand	14	0.7	1	0.2
	NA	1	0.1	0	0
Stretching in relation to a running session*	Yes	1153	58.6	333	65.6
	No	811	41.2	175	34.4
	NA	5	0.3	0	0
Wearing orthotics**	Yes	271	13.8	117	23
	No	1689	85.8	388	76.4
	NA	9	0.5	3	0.6
Participation in other sports	Yes	1448	73.5	376	74
	No	513	26.1	127	25
	NA	8	0.4	5	1

AT – Achilles tendon, NA – Not Available, * - statistically significant difference between uninjured and injured runners (p<0.05), ** - statistically significant difference between uninjured and injured runners (p<0.001).

3.3.5 Logistic regression model of the factors associated with Achilles tendon injuries

Simple logistic regression analysis identified seven significant variables ($p \leq 0.1$): sex ($p<0.001$), age group ($p<0.001$), years of running experience ($p<0.001$), race pace ($p<0.001$), running terrain (p 0.056), stretching (p 0.005) and wearing orthotics ($p<0.001$). These variables were included in the multiple logistic regression analysis to identify those variables, which would be statistically significant, and statistically associated with an injury status, at the presence of other variables. As a result, the developed logistic regression model comprised six variables as the variable, which described preferred running terrain, was non-significant when grouped with other factors (Table 3.7). Male runners were twice more likely to report Achilles tendon injuries than female runners. Runners aged between 35 and 50 years were 70 more likely to report Achilles tendon injuries in comparison to younger runners aged less than 35 years. Similarly, older runners aged over 50 years were 46 more likely to report Achilles tendon injuries comparing to the reference group of young runners. Runners with one year or less of running experience were 59 less likely to report running-related injuries than runners with ten and more years of running practice. The most commonly reported race pace of 5-6 min/km was selected as a reference category for this variable. Runners who reported their typical race pace faster than 4 min/km were 1.5 times more likely to report Achilles tendon injuries. Runners who incorporated stretching in their running routine were 41 more likely to report Achilles tendon injuries. Finally, runners who reported wearing orthotics were almost twice more likely to report Achilles tendon injuries.

Table 3.7 Logistic regression model of independent factors associated with the risk of Achilles tendon injuries.

Variable	OR (95% CI)	*p*-value
Sex (0=Female, 1=Male)	2.01 (1.6 to 2.52)	<0.001
Age group		
<35 years (reference category)	1.00	
35-50 years	1.7 (1.31 to 2.21)	<0.001
50 years	1.46 (1.05 to 2.04)	0.03
Years of running experience		
≤1 year	0.41 (0.26 to 0.64)	<0.001
2 years	0.72 (0.48 to 1.07)	0.1
3-5 years	0.88 (0.68 to 1.15)	0.36
6-9 years	1.06 (0.79 to 1.43)	0.7
≥10 years (reference category)	1.00	
Race pace		
<4 min/km	1.51 (1.04 to 2.2)	0.03
4-5 min/km	1.2 (0.93 to 1.54)	0.16
5-6 min/km (reference category)	1.00	
6-7 min/km	1.06 (0.76 to 1.49)	0.72
7 min/km	1.09 (0.59 to 2.04)	0.78
Stretching in relation to a running session (0=No, 1=Yes)	1.41 (1.15 to 1.75)	0.001
Wearing orthotics (0=No, 1=Yes)	1.93 (1.5 to 2.5)	<0.001

OR – odds ratio; CI – confidence interval; Hosmer-Lemeshow test χ^2 12.08, *p* 0.15.

3.3.6 Factors associated with bone stress injuries

Recreational runners who reported at least one case of a bone stress injury occurred in the past two years (n 475) were merged with uninjured runners (n 1969) into the BS injury cohort (N 2,444). The injured and uninjured subgroups were statistically compared across physical and training characteristics (Table 3.8; Table 3.9). Rates of male and female runners in the injured and uninjured subgroups were similar (χ^2_1 0.82, p 0.29). The compared subgroups were statistically different by the age group (χ^2_2 34.61, p<0.001) and BMI group distributions (χ^2_2 10.2, p 0.02). Although the subgroups were similar in the reported numbers of running sessions per week (χ^2_2 1.54, p 0.46), they statistically differed across the remaining training characteristics. Thus, there were significantly more runners running over 40 km per week in the injured subgroup than in the uninjured subgroup (30.9 versus 24.6%, χ^2_2 10.11, p 0.006). The subgroups also significantly differed by years of running experience (χ^2_4 26.03, p<0.001), race pace (χ^2_4 13.59, p 0.009) and preferred running terrain (χ^2_6 22.78, p 0.001). Additionally, it was more typical for injured runners than uninjured runners to report the incorporation of stretching into running practice (67.6 versus 58.6%, χ^2_1 12.98, p<0.001) and wearing orthotics (29.5% versus 13.8%, χ^2_1 68.32, p<0.001). However, the subgroups had very similar rates of the reported participation in other sports (χ^2_1 0.12, p 0.73).

Table 3.8 Physical characteristics of uninjured runners and runners with bone stress injuries.

Physical characteristics		Uninjured runners (*N*=1969)		BS Injured runners (*N*=475)	
		n	**%**	***n***	**%**
Sex	Male	866	44	198	41.7
	Female	1103	56	277	58.3
Age groups**	18-34 years	600	30.5	193	40.6
	35-50 years	982	49.9	237	49.9
	50 years	387	19.7	45	9.5
Body mass index (BMI) groups*	Normal	1465	74.4	354	74.5
	Underweight	57	2.9	17	3.6
	Overweight	405	20.6	83	17.5
	Obese	42	2.1	21	4.4

BS – bone stress; * - statistically significant difference between uninjured and injured runners (p<0.05), ** - statistically significant difference between uninjured and injured runners (p<0.001).

Table 3.9 Training characteristics of uninjured runners and runners with bone stress injuries.

Training characteristics		Uninjured runners (N=1969)		BS Injured runners (N=475)	
		n	**%**	***n***	**%**
Run sessions per week	1 or 3 sessions	852	43.3	193	40.9
	4 or 5 sessions	897	45.6	231	48.9
	6+ sessions	211	10.7	48	10.2
	NA	9	0.5	9	0.5
Weekly running distance*	15-20 km	559	28.4	109	22.9
	20-40 km	925	47	219	46.1
	40+ km	485	24.6	147	30.9
Running experience**	≤ 1 year	252	12.8	29	6.1
	2 years	197	10	55	11.6
	3-5 years	505	25.6	146	30.7
	6-9 years	280	14.2	88	18.5
	10+ years	732	37.2	157	33.1
	NA	3	0.2	0	0
Race pace*	<4 min/km	163	8.3	54	11.4
	4-5 min/km	640	32.5	179	37.8
	5-6 min/km	781	39.7	173	36.5
	6-7 min/km	306	15.5	53	11.2
	7 min km	74	3.8	15	3.2
	NA	5	0.3	0	0
Running terrain*	Bitumen	883	44.8	211	44.4
	Cement	628	31.9	133	28
	Hard dirt/gravel	342	17.4	94	19.8
	Grass	42	2.1	23	4.8
	Treadmill	53	2.7	9	1.9
	Synthetic	6	0.3	5	1.1
	Sand	14	0.7	0	0
	NA	1	0.1	0	0
Stretching in relation to a running session**	Yes	1153	58.6	321	67.6
	No	811	41.2	153	32.2
	NA	5	0.3	1	0.2
Wearing orthotics**	Yes	271	13.8	140	29.5
	No	1689	85.8	331	69.7
	NA	9	0.5	4	0.8
Participation in other sports	Yes	1448	73.5	353	74.3
	No	513	26.1	120	25.3
	NA	8	0.4	2	0.4

BS – bone stress, NA – Not Available, * - statistically significant difference between uninjured and injured runners (p<0.05), ** - statistically significant difference between uninjured and injured runners (p<0.001).

3.3.7 Logistic regression model of the factors associated with bone stress injuries

To identify factors associated with bone stress injuries in the presence of other factors, simple logistic regression was performed for each variable. Eight statistically significant variables were included in the logistic regression model: age group (p<0.001), BMI group (p 0.02), years of running experience (p<0.001), running km per week (p 0.007), race pace (p 0.009), running terrain (p 0.006), stretching (p<0.001) and wearing orthotics (p<0.001). The final model comprised seven factors as the variable 'race pace' became non-significant at the presence of other variables (Table 3.10). The age group of under 35 years was a reference category, and runners aged between 35 and 50 years and aged over 50 years were 25 and 65 less likely to report bone stress injuries, respectively. BMI was a significant factor for bone stress injuries – obese runners were 2.8 times more likely to report injuries than runners with normal BMI. However, underweight and overweight categories were not significantly different from the normal BMI category. In relation to the running experience, the group of runners with ten and more years of running experience was the reference category. Inexperienced runners, with one year or less of running, were 43 less likely to report bone stress injuries, whereas runners with 2-5 years and 6-9 years of running experience were 36 and 38 more likely to report this type of injuries. A longer weekly running distance was associated with an increased risk of bone stress injuries. Thus, runners who reported that they would run over 40 km per week had a 1.6 times higher risk when compared to the reference category of 15 to 20 km per week. Regarding the preferred running terrain, bitumen was the most common reported terrain and thus was selected as a reference category. Only one out of six terrains were associated with twice as higher chance of reporting bone stress injuries compared to the 'grass' reference running terrain. Additionally, stretching in relation to running and wearing orthotics were associated with 1.4 and 2.7 higher chance of reporting bone stress injuries, respectively.

Table 3.10 Logistic regression model of independent factors associated with risk of bone stress injuries.

Variable	OR (95% CI)	*p*-value
Age group		
<35 years (reference category)	1.00	
35-50 years	0.75 (0.6 to 0.95)	0.02
50 years	0.35 (0.24 to 0.5)	<0.001
BMI group		
Normal (reference category)	1.00	
Underweight	1.03 (0.56 to 1.89)	0.92
Overweight	1 (0.75 to 1.31)	0.97
Obese	2.79 (1.55 to 5.05)	0.001
Years of running experience		
≤1 year	0.57 (0.37 to 0.89)	0.01
2 years	1.35 (0.94 to 1.95)	0.11
3-5 years	1.36 (1.04 to 1.78)	0.03
6-9 years	1.38 (1.01 to 1.89)	0.045
≥10 years (reference category)	1.00	
Running km per week		
15-20 km (reference category)	1.00	
20-40 km	1.26 (0.96 to 1.64)	0.09
40 km	1.6 (1.18 to 2.16)	0.002
Running terrain		
Bitumen (reference category)	1.00	
Cement	0.87 (0.68 to 1.12)	0.29
Grass	2 (1.14 to 3.53)	0.02
Hard dirt/gravel	1.06 (0.8 to 1.41)	0.69
Sand	0.00	1.00
Synthetic	2.08 (0.59 to 7.3)	0.25
Treadmill	0.68 (0.32 to 1.45)	0.32
Stretching in relation to a running session (0=No, 1=Yes)	1.44 (1.15 to 1.8)	0.001
Wearing orthotics (0=No, 1=Yes)	2.73 (2.13 to 3.5)	<0.001

OR – odds ratio; CI – confidence interval; Hosmer-Lemeshow test χ^2 9.14, *p* 0.33.

3.4 Discussion

3.4.1 Discussion: reported running-related injuries

This study was the first to collect and report injury data from approximately 5,000 Australian recreational runners. One of the objectives of the project was to report frequencies of the running-related injuries in Australian runners. In this study, 54 of runners reported running-related injuries occurred in the past two years. Similarly, a recent cross-sectional study of 1,145 novice and recreational runners recruited through *parkrun* reported a 49.8 injury rate (Linton & Valentin, 2018). A prospective cohort study followed 191 recreational runners over a 12-week period and registered 60 (31) injured runners (Hespanhol Junior, Costa, & Lopes, 2013). Another prospective two-year study of 300 runners reported that 66 of the participants sustained at least one injury (Messier et al., 2018). A systematic review of 17 studies reported the incidence of running-related injuries varying between 19.4 and 79.3 with a follow-up period ranging between one day and 18 months (van Gent et al., 2007). Whilst the reporting period in this project' questionnaire was longer than in studies included in this systematic review, the injury rate of 54 was similar to the average rate reported in the systematic review. However, differences in the population characteristics, injury definitions and study designs were present. Our study survey was designed to not only collect epidemiological and injury data from recreational runners but to identify runners eligible for the genetic arm of the study. Therefore, the survey design and included questions also differed from other running injury studies, yet met our recruitment needs and was demonstrated to be a reliable questionnaire (Domaschenz et al., 2015).

The two most common reported injuries in this study were an Achilles tendon injury and a bone stress injury. Each type of injury was of approximately 16 of all reported running-related injuries. According to a systematic review, the same injuries were reported as the most common running-related musculoskeletal injuries, however, their rates were around 10 for Achilles tendon injuries and 13-20 for bone stress injuries (Dias Lopes et al., 2012). Importantly, the criteria of this systematic review required a specific definition of the injury, but not an injured anatomic part of the leg. Reported types of different injuries were classified and grouped by specific diagnoses and anatomic locations of these injuries. Van Gent et al. showed that the most common location of running-related injuries was the knee (van Gent et al., 2007). However, in this study, knee injuries were the fourth most common reported injuries.

The variability in recovery time was similar to a reported range of recovery time by Nielsen et al. The reported number of weeks off running after an injury occurred, showed that only 7.4 of runners returned to running after a one-week break, whereas about 30 of injuries required at least six weeks of recovery time. In comparison, a study of novice runners showed that the median recovery time among 254 runners was 71 days (10 weeks), varying between 9 and 617 days (Nielsen et al., 2014).

Overall, our study demonstrated high heterogeneity of sustained running-related injuries, as well as variability in the recovery time. It is important to acknowledge that the two most reported injuries were also injuries of interest for the subsequent genetic analyses – Achilles tendon injuries and bone stress injuries. Although the description of the study available for public and promotion material included information about the scope of study and its focus on genetic predisposition to tendon and bone injuries, the recruitment process was not biased towards the search of runners with these two types of injuries. The study aimed to collect an extensive amount of injury data in Australian physically active community of recreational runners both injured and uninjured and was promoted as such.

3.4.2 Discussion: risk factors associated with running-related injuries, and specifically Achilles tendon injuries and bone stress injuries

Recreational running has multiple health benefits as it is easily accessible. However, a risk of running-related injuries affects the health of runners and regularity of their physical activity. Therefore, prevention of running-related injuries is an important issue for the public and medical system. Throughout decades of research into risk factors contributing to running-related injuries, the findings have been relatively heterogeneous due to variation in data collection and analyses. Several factors were summarised in a meta-analysis as important contributors to the risk of injuries: longer weekly running distance for male runners, history of the previous injury in the past 12 months (Saragiotto et al., 2014). Additionally, an increase in training distance per week was protective for knee injuries (van Gent et al., 2007). Our study analysed the largest cohort of Australian recreational runners and attempted to contribute to the knowledge of risk factors of running-related injuries identified in the studied cohort.

Statistical analysis of data from 4,284 recreational runners demonstrated that there were significantly more male runners in the injured subgroup. The following regression analysis

showed that male sex was associated with a higher chance of reporting running-related injuries. Sex was assessed as a potential factor in multiple studies, however, the majority of the studies, which investigated overall injury risk, did not find any associations with sex (Macera et al., 1989; van Gent et al., 2007; Walter et al., 1989). However, one study found an association between male sex and higher rates of running-related injuries among younger runners aged under 40 years (McKean, Manson, & Stanish, 2006). Additionally, Linton et al. reported that among recreational runners recruited via *parkrun* men were 1.45 times more likely to report running-related injuries than women (Linton & Valentin, 2018). In addition to overall injury risk associated with male sex, our study demonstrated that male runners were more likely to report Achilles tendon injuries. These findings corresponded with the higher rates of Achilles tendinopathy in male runners in the study of 2002 running injuries (Taunton et al., 2002). Although men are more likely to report Achilles tendon injuries than women, previously described sex differences in training loads and patterns (Section 2.4.1), including running longer weekly distances and more frequently, may explain these differences and therefore, complicate the evaluation of sex as an independent factor. Interestingly, previous literature reviews demonstrated that bone stress injuries were more typical for female runners (Mattila et al., 2007; Newman et al., 2013). However, in our study, the frequencies of bone stress injuries did not differ between male and female runners. Similarly, to Achilles tendon injuries, sex should not be considered as an independent factor, being possibly associated with training factors and other sex-related factors.

Age was a factor associated with overall injury risk, as well as for both Achilles tendon injuries and bone stress injuries. Regression analysis showed that runners aged over 50 years were less likely to report running-related injuries than runners under 35 years of age. Previously, older age was demonstrated as a contradictory factor, which may increase risk of injuries, yet with limited evidence (van der Worp et al., 2015; van Gent et al., 2007). Potentially, our study's results could be explained by the speculation that injuries at an older age would make runners reconsider their training routine and stop running. Interestingly, different age groups were associated with reporting of Achilles tendon injuries and bone stress injuries. Thus, age groups of 35 to 50 years and over 50 years were associated with 1.7 and 1.5 increased risk of Achilles tendon injuries, respectively. These results were consistent with the literature review reporting that people aged between 30 and 55 had the highest rates of Achilles tendinopathy, which could be explained by molecular changes in the tendon, as ageing processes start in the body (Cook, Khan, & Purdam, 2002).

Conversely, the same age groups were associated with decreased risk of bone stress injuries in the BS injury cohort. Although bone density decreases with age, and reduced bone density potentially increases the risk of injuries in older age, younger runners may also be at risk of bone stress injuries due to unfinished maturation of bone and its susceptibility to injuries at that time. This could explain why runners aged under 35 years were more likely to report bone stress injuries than runners older than 35 years.

In this study, BMI was not a significant factor for running-related injuries and Achilles tendon injuries, however, was a contributing factor to the risk of bone stress injuries. Obese runners were 2.8 times more likely to report bone stress injuries when compared to runners with normal BMI. It is important to acknowledge that the BS injury cohort comprised only 63 obese runners, and one-third of these runners (n 21) reported bone stress injuries. Several previous studies also demonstrated that increased BMI was associated with the risk of bone stress injuries, such as medial tibial stress syndrome and tibial stress injuries (Beck et al., 2015; Newman et al., 2013).

The performed logistic regression analyses demonstrated that runners with only one year or less of running experience were less likely to report running-related injuries overall, as well as bone stress injuries and Achilles tendon injuries, when compared to runners with more than ten years of running practice (a reference group). However, in the survey, the respondents were asked about injuries that occurred only in the past two years. Therefore, the reported lower rates of sustained injuries among inexperienced runners could be explained by a shorter period of running practice. Interestingly, uninjured and injured subgroups comprised of similar proportions of runners with two years of running experience, and their injury rates did not differ significantly from runners with over ten years of running experience. Interestingly, runners with 3-5 and 6-9 years of running experience were more likely to report bone stress injuries than the reference group. Similar observations were reported in a systematic review of risk factors of MTSS: fewer years of running experience were associated with the development of MTSS (Newman et al., 2013). Potentially, running experience could be one of the markers of muscular and bone adaptation to the load, in conjunction with running volume, frequencies of training sessions, and running terrain. Therefore, correct preconditioning of runners and adaptation to load may reduce injury risk making this factor modifiable through training adjustments.

Two other factors characterising training characteristics were also associated with a higher chance of reporting bone stress injuries, but not other types of injuries: longer weekly

running distance and running on grass. Thus, runners who would run 20-40 km per week and over 40 km per week were 26 and 60 more likely to report bone stress injuries than runners with 15-20 km of weekly running distance, respectively. Longer distance may be a parameter of either high training load or gradual increase of the load. Both may contribute to the accumulation of microdamage in the bone and subsequent development of bone stress injuries (Warden et al., 2014). Preferred running terrain may also influence bone loading. Originally harder surfaces such as bitumen and cement were hypothesised to increase the load to bone comparing to more compliant surfaces such as grass and sand. However, no studies have identified a clear association between running terrain and the risk of bone stress injuries. Interestingly, the risk of bone stress injuries was associated with the recent change in terrain, as this change requires adaptation to different ground reaction forces and therefore, may increase the risk of microdamage in the bone (Warden et al., 2014). Therefore, we did not find any literature to support our finding of the running terrain 'grass' as a potential risk factor. However, it should be taken into consideration that only 65 runners reported 'grass' as their preferred running terrain, and one-third of those runners reported bone stress injuries.

The number of running sessions per week was a significant factor for development of running-related injuries. Runners who reported running 1-3 and 4-5 times per week were more likely to report running-related injuries than runners who would run at least six times per week. According to a systematic review, some of the previous studies reported an increased risk of injuries associated with the increase in the number of running sessions per week, whereas other studies did not find this association (van Gent et al., 2007). Although our results may differ from the previous findings, may be considered in the following research.

Race pace of less than 4 minutes per kilometre was associated with 1.5 higher chance of Achilles tendon injuries. However, this variable was not associated with either risk of all running-related injuries or bone stress injuries. McCrory et al. reported that runners with Achilles tendinopathy would train at a higher pace, prior to the injury, comparing to uninjured runners (McCrory et al., 1999). Another study demonstrated that runners competing at shorter distances of 1,500-3,000 metres were more likely to develop Achilles tendinopathy than runners competing at the 10 km race. Additionally, Achilles tendinopathy was the most common injury among elite runners in this study, as they would train at faster paces than recreational runners (Knobloch et al., 2008). Our findings support these observations and

suggest that faster race pace may be an important contributor to the risk of Achilles tendon injuries.

Two remaining factors associated with a higher chance of reporting both types of injuries and overall running-related injuries were stretching and wearing orthotics. Stretching was often considered as a part of injury prevention program as muscle stiffness, and restricted range of motion were factors for musculoskeletal injuries in runners (Van Mechelen, Hlobil, Kemper, Voorn, & de Jongh, 1993). However, a systematic review of the impact of stretching on sports injury risk concluded that included studies failed to demonstrate any protective effect of stretching, both in cohort studies and randomised controlled trials (Thacker, Gilchrist, Stroup, & Kimsey Jr, 2004). Moreover, this study suggested that not only may not prevent sports injuries but also might compromise sports performance as stretching may decrease joint stability, decrease muscle and tendon ability to absorb energy and change body position leading to overloading of tissues. Although the results of our study showed an association between the development of injuries and stretching, we were unable to derive any conclusions about their link as our study lacked details about the timeline of injuries and introduction of the stretching into running practice. Orthotics influence the biomechanics of the foot and usually are used for improving running and for additional cushioning (McKenzie et al., 1985). Therefore, correctly prescribed orthotics may assist in injury prevention as well as recovery from injuries. However, an inaccurate choice of orthotics may increase risk of injury. The results of our study showed that injured runners would wear orthotics significantly more often than uninjured runners, yet our data were missing the time link between the injury occurrence and the introduction of orthotics. Similarly, to stretching, we were unable to conclude on the role of orthotics in the injury risk, and therefore these two factors require additional data collection with more detailed questions.

In summary, our study identified factors associated with the development of running-related injuries and two most common overuse injuries – Achilles tendon injuries and bone stress injuries. These findings may contribute to the injury prevention programs, optimisation of running programs and further research of risk factors of running-related injuries.

4. Chapter Four – Genome-wide association study of genetic variants associated with Achilles tendon injuries

Addendum

Contributions to Chapter 4:

Mariia Kozlovskaia:

- Sample collection and registration
- DNA extraction
- GWAS data analysis
- Imputed data analysis
- Author of the chapter

Staff members of the Translational Research Institute:

- GWAS analysis

Paul Leo:

- Data imputation
- Assistance with GWAS data analysis

Kevin Ashton:

- Assistance with GWAS data interpretation
- Editing of the chapter

Rebecca Grealy

- Sample registration
- DNA extraction

Nicole Vlahovich

- Editing of the chapter

4.1 Introduction

An Achilles tendon injury is one of the most common lower leg injuries amongst runners. A systematic review of running-related injuries reported that a typical rate of Achilles tendon injuries was 9.1-10.9 (Dias Lopes et al., 2012). Multiple risk factors have been associated with Achilles tendon injuries. Risk factors are divided into modifiable and non-modifiable; with modifiable risk factors comprising of training, lifestyle factors and medications, which may affect tendons, and non-modifiable comprising of physical and anatomic characteristics and genetic variants associated with the development of Achilles tendon injuries.

Although training characteristics such as running distances and the number of years of running did not show consistent associations with the development of Achilles tendon injuries (Lorimer & Hume, 2014), running on soft terrains such as sand has been shown to increase the risk of Achilles tendon injuries, whereas running on asphalt was a protective factor against this type of injury (Knobloch et al., 2008). Lifestyle habits such as smoking cigarettes and alcohol consumption are also considered potential risk factors to developing Achilles tendon injuries, yet show contradictory results or very moderate associations with the development of the injuries, and could alternatively be explained by training-related factors (Owens et al., 2013). Several studies have shown a significant effect of fluoroquinolone antibiotics on the increased risk of Achilles tendon injuries, in particular, tendon ruptures, with a reported three-fold increase in a Danish cohort and 4.1-fold increase in an Italian cohort (Corrao et al., 2006; Sode et al., 2007). Moreover, with additional exposure to corticosteroids risk of tendon rupture increased to 10-fold in the same Italian population (Corrao et al., 2006).

It is more typical for men to report Achilles tendon injuries, yet this difference could be explained by differences in training loads between sexes (Magnan et al., 2014). A systematic review regarding the pathogenesis of Achilles tendinopathy showed that age was a significant factor affecting the tendon matrix (Magnan et al., 2014). With the highest rate of Achilles tendinopathy observed amongst individuals aged between 30 and 55 (Cook et al., 2002). A cohort study of long-distance runners also showed a significant association of older age and risk of Achilles tendinopathy (Hirschm ller et al., 2012). Specific anatomic characteristics, such as malalignment and hyperpronation, were shown to be associated with increased risk of Achilles tendinopathy (J rvinen et al., 2005). At the same time, correction with orthotics of foot anatomy may prevent the risk of Achilles tendon injuries

(Peters et al., 2015). Stretching may also contribute to the Achilles tendon injury, although the evidence of this is currently inconclusive (Peters et al., 2015).

The majority of genetic studies investigating the risk of Achilles tendinopathy have employed a candidate gene approach, targeting genes which encode different types of collagen and other molecules responsible for the regulation of extracellular matrix homeostasis. The most investigated gene was *COL5A1*, with several genetic variants in this gene being associated with the development of Achilles tendon injuries across three Caucasian populations: Australian, South-African and British (Brown et al., 2016; Mokone et al., 2006; September, 2009). Genetic variation in the *TNC* gene, a regulator of cell-matrix interaction in tendon tissue, was associated with decreased risk of Achilles tendinopathy (Mokone et al., 2005). Further investigation of proteins involved in the tendon structure has also targeted fibrillin and elastin, due to their role in tendon elasticity, strength and flexibility. However, only a single association between a variant in the *FBN2* gene and injury has been shown in Australian and South-African cohorts (El Khoury et al., 2015). Metalloproteinase 3 (MMP3) is a key regulator of extracellular matrix homeostasis, and three investigated polymorphisms in the *MMP3* gene were significantly associated with decreased risk of Achilles tendinopathy (Raleigh, 2009). Tissue inhibitors of metalloproteinases (TIMP) have also been considered as potential contributors to the genetic risk of Achilles tendinopathy. Genetic variation in the *TIMP2* gene was significantly associated with reduced risk of Achilles tendinopathy in two studies, which concluded that the balance between metalloproteinase and TIMPs function may be important for the development of Achilles tendinopathy (El Khoury et al., 2013; El Khoury et al., 2016). Since several growth factors play an important role in tendon growth and homeostasis, one study investigated the role of polymorphisms in the genes *TGFB1* and *GDF5*, which encode transforming growth factor-β1 and growth/differentiation factor-5, respectively. This study found a significant association of one polymorphism located in *GDF5* but did not find any association between *TGFB1* and Achilles tendinopathy (Posthumus et al., 2010). Caspases are involved in the pathways accompanying tendon cell apoptosis, and their expression was elevated in tendinopathy cases. A genetic polymorphism located in the *CASP8* gene has been associated with the risk of Achilles tendinopathy in Australian and South-African populations (Nell et al., 2012). Although a number of reviewed genes were associated with development of Achilles tendon injuries, multiple other genes with relevant roles in tendon structure (*COL11A1, COL12A1, COL14A1, COL27A1*) and tendon homeostasis (*NOS2, NOS3, CASP3*) did not show any

association with the condition (Hay et al., 2013; Nell et al., 2012; Rickaby et al., 2015; September et al., 2008).

Since Achilles tendinopathy is a multifactorial condition, and the contribution of each polymorphism is relatively small, it is important to acknowledge the importance of the sufficient samples size and unbiased approach when candidate genes are investigated. Unfortunately, the majority of these studies incorporated a candidate gene approach and used relatively small cohort size, with approximately 150 cases and 350 controls in the largest studied cohorts, which could lead to biased results and hence require replication analysis, which could prove their findings.

With the development of high-density molecular methods, it is now possible to investigate thousands-to-millions of genetic variants at one time, at relatively low cost. GWAS is one such approach and based upon the principle of linkage disequilibrium (LD) at the population level. As a GWAS enables the genotyping of thousands-to-millions of genetic markers, simultaneously across the whole genome, it avoids the bias of preselection of potentially important genes as observed in the candidate gene approach (Hartl et al., 1997; Visscher, Brown, McCarthy, & Yang, 2012). This unbiased genome-wide approach has been used in the search for single-nucleotide polymorphisms (SNPs) associated with many complex diseases (Donnelly, 2011). Within the first five years of its introduction, GWAS identified many new associated genes in relation to a number of autoimmune diseases such as ankylosing spondylitis, rheumatoid arthritis and type 1 diabetes, as well as metabolic diseases such as type 2 diabetes and obesity (Visscher et al., 2012). However, one of the critical considerations in GWAS is that since each polymorphism may contribute a small effect on the studied trait, it is essential to have a large sample size to provide statistically reliable results.

To date, only a single GWAS has investigated genetic variants associated with Achilles tendon injuries (Kim et al., 2017). This study used genomic data of a cohort, which comprised of over 100,000 hospital patients, including over 5,000 patients with Achilles tendon injuries. Whilst no statistically significant SNPs were identified, four SNPs reached a suggestive significance level (1×10^{-6}). Two of these four SNPs (rs57104447, rs60713544) were located in the intragenic regions on chromosomes 1 and 4, respectively. The other two identified SNPs (rs1937810, rs57224706) were located in the *MMP7* and *SMARCD1* genes, respectively. In addition, previously identified SNPs from candidate gene studies were unable to be replicated in this cohort (Kim et al., 2017). The genetic contribution to the

Achilles tendon injury predisposition is still unclear and requires further investigation, therefore this part of the project aimed to identify novel polymorphisms associated with Achilles tendon injuries in recreational runners. Also, results from this cohort of recreational runners were assessed for SNPs previously associated with Achilles tendinopathy predisposition and protection.

4.2 Methods

4.2.1 Participants

Participants for the genomic study were a subset of 1,165 (23.6) recreational runners derived from 4,720 respondents (described in Chapter 2) who met the following eligibility criteria:

- 18-50 years of age,
- reported background of at least 75 of Caucasian European or Mediterranean
- current non-smoker with at least five years after giving up,
- have no history of chronic conditions that could affect the musculoskeletal system (osteoarthritis, osteoporosis, rheumatoid arthritis, chronic renal failure),
- no history of chemotherapy.

With the application of these criteria, only runners who reported injuries, which were more likely to be running-related, but not affected by ageing, chronic conditions, smoking and chemotherapy. In addition, those runners who reported Achilles tendon injuries and bone stress injuries were included in the genetic study if they did not have a fracture or surgery preceding the injury, and in case of only an Achilles tendon injury did not report corticosteroid injections and the use of quinolone antibiotics in the past six months.

4.2.2 Sample collection

Eligible participants were contacted via email to confirm their agreement to provide a saliva sample. In these emails, eligible participants also received two documents: 'Participant information sheet' and 'The ethics of genetic research', aiming to provide an understanding of the project, and the possible benefits and risks of participating in genetic research (Appendices 8 and 9). Saliva as a source of genomic DNA was chosen due to the non-invasive, simple collection method and the ability to transport these samples at room temperature through the mail. Oragene DNA (OG-500) collection kits (DNA_Genotek, 2015) were selected for the sample collection due to simple instructions to provide a saliva sample, long-term sample storage at room temperature and expected high median DNA yield (110ng/2mL of saliva). Participants were mailed the collection kit, all associated paperwork, and a reply-paid envelope and were asked to return the sample and consent forms via the

mail (Appendix 10). Samples were received at Bond University and registered in BC|SNP system (BC, 2015). Saliva samples were stored at room temperature until processing.

4.2.3 DNA Extraction

DNA extractions were conducted in the Molecular Biology research laboratory at the Faculty of Health Sciences and Medicine, Bond University. DNA was extracted from saliva samples according to standard protocols with minor modifications (DNA_Genotek, 2015). Specifically, 1mL of Oragene saliva samples were incubated at 50^{o}C overnight to maximise DNA yield and inactivate nucleases. Following incubation, 40 L (1/25th volume) of PrepIT-2LP lysis solution was added to the saliva samples and mixed by vortexing. Then samples were incubated on ice for 10 minutes, then centrifuged at 4,500 rpm for 10 minutes. The supernatant (1mL) was transferred to a new tube with 1.2 mL (1.2x volume) of 100 ethanol added to precipitate the DNA. Samples were re-centrifuged to pellet the DNA precipitate and supernatant was discarded. The DNA pellet was washed with 70 ethanol (1mL), centrifuged as before, the supernatant removed, and the DNA pellet allowed to dry to remove ethanol traces. The DNA pellet was rehydrated with 200 L of ddH_2O and incubated at 50 C for one hour to ensure complete resuspension. To assess concentration and quality of the samples, DNA samples were analysed using UV spectrophotometry (NanoDrop) and fluorometry (Qubit 3). Extracted DNA was stored at -20 C prior to subsequent analysis.

4.2.4 Genotyping and Quality Control

All samples were genotyped using Infinium CoreExome-24 BeadChips (Illumina, 2015). This microarray contains 550,000 genetic markers, made up of exonic variants, splice variants, stop altering variants, ancestry informative markers and MHC tag SNPs (Illumina, 2015). All arrays were processed according to standard protocols and workflows on the Illumina iScan platform at the Australian Translational Genomics Centre, Queensland University of Technology/Translational Research Institute. Samples were processed in three batches across the two-year project timeline. Sample quality was assessed using Illumina Genome Studio v. 2011.1, and final data were released as .map and .ped files. Files were imported into the BC|SNP database (BC, 2015) and analysed. Quality Control (QC) analysis was conducted using BC Platforms software (BC, 2015), with QC thresholds set to exclude individuals with 5 failed genotypes. In addition, an X chromosome inbreeding estimate

was applied to identify sex non-concordance of genotyped samples. Samples displaying heterozygosity rates outside the range of three standard deviations of the mean were also excluded. Identity by state (IBS) was used to identify cryptic relatedness and individuals excluded that were 10 related. Finally, Principal Component Analysis (PCA) was performed to assess the population stratification of our participants. SNP genotypes from 51 populations from the Human Genome Diversity Project (HGDP) were used as our population reference (Cavalli-Sforza, 2005). HGDP genotype data and sample descriptions are freely available from: http://hagsc.org/hgdp/files.html. Thresholds for PCA analysis were set up at six standard deviations from the mean. Markers with 7.5 genotypes missing, out of Hardy-Weinberg Equilibrium (p<0.0001) and a minor allele frequency (MAF) <0.01 were excluded.

4.2.5 Statistical Analyses

After the QC steps, samples were assigned to one of two case groups (an Achilles tendon injury or a bone stress injury) or an uninjured control group. Genotyped AT genomic and BS genomic cohorts were statistically compared to the previously described AT injury cohort and BS injury cohort, respectively, using a χ^2 test. GWAS data were tested using logistic regression with an additive genetic model implemented in PLINK (Purcell et al., 2007). Genome-wide significance and suggestive levels of $p<5x10^{-8}$ and $p<1x10^{-5}$, respectively, were chosen initially, according to standard guidelines. Manhattan plots were generated using HaploView software (Barrett, Fry, Maller, & Daly, 2004), PCA plots and QQ-plots were generated in R-software (R Development Core Team, 2008). Power calculations were performed using the Genetic Power Calculator (Purcell, Cherny, & Sham, 2003).

4.2.6 Imputation

SHAPEIT2 was used for pre-phasing of the data and Sanger Imputation Service platform was used for imputation based on Haplotype Reference Consortium as a reference dataset (Delaneau, agury, & Marchini, 2013; Haplotype Reference Consortium, 2016; Sanger Institute, 2017). The outcome .vcf files containing imputed genotypes were then filtered to exclude imputed SNPs with an INFO score of <0.6 (Danecek et al., 2011). Filtered files were transformed into PLINK format (.bed, .bim, .fam) prior to data analysis. Case-control

analyses were implemented using PLINK, with the outcome association files split by chromosome and plots generated for each chromosome in Haploview (Barrett et al., 2004). The Locus oom online service was used to visualise genotyped and imputed significant SNPs (Pruim et al., 2010).

4.2.7 Replication Analysis

Using the imputed dataset, *p*-values of identified significant SNPs in this study were compared to the *p*-values of the same SNPs acquired from publicly available data from GWAS of Achilles tendon injuries by Kim et al. (Kim et al., 2017). Previously identified by candidate gene approach studies SNPs, described in Table 1.1 in Chapter 1, were further investigated using imputed data to attempt replication of those results in the current study and seeking identified significance levels for previously described SNPs in the imputed data.

4.3 Results

4.3.1 Phenotypic characteristics of the participants

4.3.1.1 Definition of investigated cohorts

Initial recruitment for the project resulted in 4,720 unique responses from recreational runners who reported running at least 15 km per week. However, in order to investigate training-related risk factors of running-related injuries, particularly Achilles tendon injuries and bone stress injuries, these responses were filtered by the circumstances the injuries occurred. Only runners who reported injuries that occurred while running (as opposed to injuries which occurred during other activities) were included in the study. Firstly, runners who reported injuries that occurred while not running were excluded (*n* 436), resulting in a cohort of 4,284 runners, corresponding to 2,315 (54) injured and 1,969 (46) uninjured runners. Secondly, individuals who reported specifically Achilles tendon injuries and bone stress injuries were grouped separately, resulting in 508 Achilles tendon injury cases and 475 bone stress injury cases. Finally, each case group was merged with the uninjured runners (*n* 1,969) for the subsequent case-control analyses. This resulted in two cohorts: the Achilles tendon (AT) injury cohort and the bone stress (BS) injury cohort. With the AT cohort comprising of 2,477 runners (508 Achilles tendon injury cases and 1,969 uninjured controls) and the BS injury cohort comprising of 2,444 runners (475 bones stress injury cases and 1,969 uninjured controls).

Selection criteria for the genetic arm of the study applied to 4,720 participants, identified 1,651 recreational runners (35) eligible for genetic analysis. These study participants were contacted for recruitment into the genomic study, of which 1,165 (70.6) provided signed consent and a saliva sample (Figure 4.1). In order to assess whether responsiveness to participate was biased by sex, age or both, a χ^2 test was performed. As shown in Table 4.1, response rates from both males and females were very similar. However, participant age was associated with willingness to participate in the study, as younger age was associated with a lower rate of response to the invitation email ($p<0.001$). Considering relatively low rates of young runners in the cohort, this association predicted higher proportions of middle-age participants in the genotyped cohort.

Table 4.1 Summary of response rates to the invitation email by sex and age group categories.

Categories		Eligible and replied with contact details	
		Yes (%)	No (%)
Sex	Male	78.4	21.6
	Female	81.4	18.6
Age group*	18-24	65	34.9
	25-34	77.3	22.7
	35-44	82.8	17.2
	45-50	82.6	17.4

*- statistically significant difference between 'Yes' and 'No' responses (p<0.001).

Of the 1,165 collected saliva samples, 1,099 passed quality control analysis and were grouped as uninjured or with a specific type of injury. This resulted in two genomic cohorts: AT genomic cohort of 938 runners (171 Achilles tendon injury cases and 767 uninjured controls) and BS genomic cohort of 941 runners (174 bone stress injury cases and 767 uninjured controls). These two cohorts were used for subsequent comparison to the AT and BS injury cohorts, as described earlier (Chapter 3). The remainder of this chapter is focused on the AT injury and genomic cohorts, while the BS cohorts are further discussed in Chapter 5.

4.3.1.2 Statistical comparison of AT genomic cohort and AT injury cohort

The physical characteristics and reported ethnic background of the participants included in the AT genomic cohort are displayed in Table 4.2, with the physical characteristics of AT injury cohort are previously presented in Table 3.5.

Within the AT genomic cohort, 18.2 were injured cases. This rate was very similar to the rate of 20.5% of injured runners in the AT injury cohort (χ^2_1 2.217, p 0.09). Whilst the AT genomic cohort had significantly higher proportion of females than the AT injury cohort (56.4 versus 52.1% χ^2_1 4.95, p 0.015), the frequencies of males and females did not significantly differ when injured and uninjured groups of the cohorts were compared to each

other (χ^2_1 2.259, p=0.08, χ^2_1 2.346, p 0.08, respectively). In regard to BMI group distribution, the AT genomic cohort had a significantly higher rate of runners with a BMI within a normal range relative to AT injury cohort (78.3% versus 73.7, χ^2_1 7.45, p 0.004). Hence, when injured and uninjured groups of AT genomic cohort were compared to injured and uninjured groups from the AT injury cohort, frequencies of normal BMI groups also differed statistically (80.7% versus 71.1%, χ^2_1 6.1, p=0.008, 77.7% versus 74.4%, χ^2_1 3.23, p 0.044).

A participant age between 18 and 50 years old was a key selection criterion for inclusion in the genomic study. However, one-fifth (20.2) of the AT injury cohort were older than 50 years of age. In order to compare the age groups between the two AT cohorts, older recreational runners were removed from the AT injury cohort, and the remaining runners categorised by the following age groups: 18-24, 25-34, 35-44 and 45-50. The same grouping was also applied to the AT genomic cohort. The distribution of age groups differed significantly between the AT injury and AT genomic cohorts (χ^2_3 12.26, p 0.007). However, the frequencies of age groups did not differ significantly between injured groups (χ^2_3 1.62, p 0.65). Only the uninjured AT genomic cohort group differed significantly from the uninjured group in the AT injury cohort by age group distribution (χ^2_3 12.23, p 0.007). This was due to the uninjured group from AT genomic cohort having significantly less young (18-24 years old) runners than the AT injury cohort (4.9 versus 7.9). This could be explained by the previously described unwillingness of younger respondents to get involved in the genetic arm of the study.

Reported ethnicity was a selection criterion for the genetic analysis. Runners reported the ethnic background of their four grandparents, which allowed for the identification of eligible participants with at least 75 of Caucasian European or Mediterranean background. Of the 2,477 runners included in the AT injury cohort, 83.7 were reportedly 100 Caucasian European and 1.4 were of 100 Mediterranean ancestry. Among runners included in the AT genomic cohort and those who passed GWAS population stratification analysis, the majority (94.8) were reportedly 100 Caucasian European, and only 5.2 of participants reported either various proportions of Caucasian European, Mediterranean and other backgrounds, or 100 Mediterranean background.

The AT genomic cohort was compared to the AT injury cohort across training characteristics, these compared cohorts did not differ by years of running experience (χ^2_4 7.66, p 0.105), number of running sessions per week (χ^2_3 1.28, p 0.74), running terrain preferences

(χ^2_4 2.81, p=0.59), stretching in relation to a running session (χ^2_1 0.103, p 0.75), wearing orthotics (χ^2_1 0.001, p 0.97) and participation in other sports besides running (χ^2_1 2.28, p=0.13). However, a significant difference in the rates of reported race pace (χ^2_4 21.89, p<0.001) was observed. Thus, the injured group from the AT genomic cohort had a significantly higher rate of the <4 min/km race pace than the injured group from the AT injury cohort (20.5% versus 9.2%, χ^2_4 27.5, p<0.001). The same significantly higher rates of faster race paces were observed in the uninjured group from the AT genomic cohort comparing to the uninjured group from the AT injury cohort (χ^2_4 13.34, p 0.01). In summary, AT genomic cohort differed from AT injury cohort by frequencies of males and females, frequencies of runners with normal BMI, frequencies of runners aged between 18-24 years old, and reported race pace.

4.3.1.3 Statistical comparison of injured and uninjured runners within the AT genomic cohort

The AT genomic cohort consisted of 410 (43.7) males and 528 (56.3) females. Within the AT genomic cohort, a higher proportion of males (n 97, 23.7) reported Achilles tendon injuries compared to females (n 74, 14). Injured and uninjured groups within the AT genomic cohort were statistically different by the proportions of males and females (χ^2_1 14.396, p<0.001). These groups also statistically differed by frequencies of age groups with significantly more runners aged 45-50 in the injured group than in the uninjured group (χ^2_3 8.895, p 0.031). Additionally, the median age of cases was 41 years, in comparison to 39 years of age for the median age of uninjured controls. This could be explained by generally higher rates of injured than uninjured runners aged 45-50 in the AT injury cohort (21.9 versus 15.4). The majority of runners, in both studied groups, those with Achilles tendon injuries and uninjured, were in the normal BMI range.

Table 4.2 Physical characteristics of runners with Achilles tendon injuries and uninjured runners in the AT genomic cohort.

Physical characteristics		Runners with AT injuries (*N*=171)		Uninjured Runners (*N*=767)	
		n	***%***	***n***	***%***
Sex**	Male	97	56.7	313	40.8
	Female	74	43.3	454	59.2
Age group*	18-24 years	5	2.9	38	4.9
	25-34 years	37	21.6	204	26.6
	35-44 years	78	45.6	371	48.4
	45-50 years	51	29.8	154	20.1
Body Mass Index	Underweight (<18.5 kg/m^2)	2	1.2	16	2.1
	Normal (18.5 to <25 kg/m^2)	138	80.7	596	77.7
	Overweight (25 to <30 kg/m^2)	26	15.2	141	18.4
	Obese (≥30 kg/m2)	5	2.9	14	1.8
Reported ethnic background	100 CE	161	94.2	728	94.9
	75 CE + 25 MT/Other	7	4.1	14	1.8
	50 CE + 50 MT/Other	1	0.6	19	2.5
	100 MT	2	1.2	5	0.7
	75 MT + 25 Other	0	0	1	0.1

ATI – Achilles tendon injury, CE – Caucasian European, MT – Mediterranean, * - statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.05), **- statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.001).

Training characteristics of the different sub-groups in the AT genomic cohort are displayed in Table 4.3. Only three parameters were statistically different: 1) years of running experience (χ^2_3 29.325, p<0.001), 2) race pace (χ^2_5 16.014, p 0.007) and 3) wearing orthotics (χ^2_2 8.244, p 0.016). Almost half (47.4) of the participants included in the case group reported more than 10 years of running experience, whereas only 35.5 of uninjured controls had over 10 years of running practice. At the same time, runners with Achilles tendon injury history were more likely to run at a race pace of less than 4 min/km compared to uninjured runners (20.5 versus 10.7). It was also more typical for the runners with Achilles tendon injuries than uninjured runners to wear orthotics (22.8 versus 14.1).

In summary, injured and uninjured groups within the AT genomic cohort differed by frequencies of males and females, frequencies of the age groups, running experience, race pace and wearing orthotics.

Table 4.3 Training characteristics of runners with Achilles tendon injuries and uninjured runners in the AT genomic cohort.

Training characteristics		Runners with AT injuries (*N*=171)		Uninjured Runners (*N*=767)	
		n	***%***	***n***	***%***
Weekly running distance	15-20	36	21.1	205	26.7
	20-40	83	48.5	362	47.2
	40+	52	30.4	200	26.1
Running experience**	≤ 1 year	6	3.5	99	12.9
	2 years	9	5.3	73	9.5
	3-5 years	30	17.5	196	25.5
	6-9 years	45	26.3	127	16.5
	10+ years	81	47.4	272	35.5
Running sessions per week	1 session	0	0.0	2	0.3
	2 or 3 sessions	75	43.9	327	42.6
	4 or 5 sessions	83	48.5	358	46.7
	6+ sessions	13	7.6	74	9.6
	N/A	0	0.0	6	0.8
Race pace*	<4 min/km	35	20.5	82	10.7
	4-5 min/km	67	39.2	278	36.2
	5-6 min/km	51	29.8	291	37.9
	6-7 min/km	15	8.8	95	12.4
	7 min/km	3	1.8	18	2.3
	N/A	0	0.0	3	0.4
Running terrain	Bitumen	79	46.2	322	42.0
	Cement	42	24.6	256	33.4
	Grass	3	1.8	16	2.1
	Hard dirt/gravel	43	25.1	146	19.0
	Treadmill	4	2.3	20	2.6
	Sand	0	0.0	4	0.5
	Synthetic	0	0.0	3	0.4

Table 4.3. Training characteristics of runners with Achilles tendon injuries and uninjured runners in the AT genomic cohort (continuation).

Training characteristics		Runners with ATI (*N*=171)		Uninjured Runners (*N*=767)	
		n	***%***	***n***	***%***
Participation in other sports	Yes	132	77.2	583	76.0
	No	37	21.6	182	23.7
	N/A	0	0.0	2	0.3
Stretching in association with a running session	Yes	110	64.3	447	58.3
	No	61	35.7	318	41.5
	N/A	0	0.0	2	0.3
Wearing orthotics*	Yes	39	22.8	108	14.1
	No	132	77.2	658	85.8
	N/A	0	0.0	1	0.1

* - statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.05), **- statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.001).

4.3.1.4 Discussion of phenotypic characteristics of investigated cohorts

Here we present data showing that both the AT injury cohort and AT genomic cohort comprised more females than males. However, the rate of the Achilles tendon injury was higher in male runners than in female runners in both cohorts. Higher prevalence of female participants in the cohorts could be explained due to the use of social media as the key recruitment strategy (Manzanero, Kozlovskaia, Vlahovich, & Hughes, 2018), as well as the conclusions from a systematic review showing that women were more likely to participate in survey research if promoted through social media (Thornton et al., 2016). Previously it has been shown that Achilles tendon injuries are more typical for males (Maffulli, Wong, & Almekinders, 2003). An analysis of running-related injuries in 2,002 runners identified a significantly higher rate of Achilles tendinopathy in male runners compared to female runners (Taunton et al., 2002). Taunton et al. (2002) also identified that being aged under 34 years is a protective factor against Achilles tendinopathy for male runners. Another study showed that older age is a risk factor of Achilles tendinopathy (Hirschm ller et al., 2012). Therefore, the findings in our cohorts that the majority of injured runners were aged between 35 and 50 years in both AT genomic and AT injury cohorts supports previous findings.

However, the higher rate of middle age runners participated in the genetic analysis also could be explained by the finding that younger runners were more likely to reject the invitation to the genetic arm of the study (Manzanero et al., 2018). However, this recruitment bias affected only the uninjured group of the AT genomic cohort. Interestingly, this avoidance of participation in the genetic research by younger people was identified in other health-related studies, which required the collection of DNA samples (McQuillan & Porter, 2011; Mezuk, Eaton, & andi, 2008).

4.3.2 Sample and genotype data quality control

Of the 4,720 responses, 1,651 (35) recreational runners were eligible for the genetic arm of the study, being either uninjured or reporting an Achilles tendon injury or a bone stress injury, or both. These eligible runners were contacted via email with a request to provide a saliva sample. Of those contacted, 1,315 (79.6 eligible individuals) runners agreed to provide a saliva sample. A final total of 1,165 runners (88.6 return rate) submitted saliva samples for genetic analysis (Figure 4.1). In addition, 35 participants were recontacted to provide a second saliva sample due to low DNA yields.

Quality control steps were performed on all participants. Firstly, 18 samples (1.6) were excluded due to genotyping call rates of <95 or sex non-concordance. Additional sample QC steps identified 48 (4.1) samples that were outside of the calculated thresholds for missingness (*n* 17), heterozygosity (*n* 4), relatedness (*n* 17) and population stratification analysis for ethnicity (*n* 10). As shown in Figure 4.2, the majority of the participants were within thresholds, matching the reported ethnic background of at least 75 of being Caucasian European or Mediterranean. There were ten outliers who had another dominating ethnic background. Three runners were allocated to Druze (Arabic) ethnic group, one runner was close to Burusho (Pakistani) ethnic group, and the remaining six were outside ethnicity thresholds towards Asian ethnic groups.

Following sample QC, 1,099 participants passed and were included in the case-control analysis. Of these participants, 171 (15.6) had a previous Achilles tendon injury (AT genomic cohort), 174 (15.8) had a previous bone stress injury (BS genomic cohort), 13 individuals were included in both groups as they reported both types of injuries, and 767 (69.8) uninjured controls (Figure 4.1).

A total of 535,190 genetic variants were initially genotyped in the case-control analysis, QC filtering, however, removed 251,626 variants due to a MAF threshold of <1 , a further 2,145 variants were removed due to missing genotyped data and 251 variants failed Hardy-Weinberg exact test. Following QC filtering a total of 281,168 genotyped genetic variants remained for association analysis.

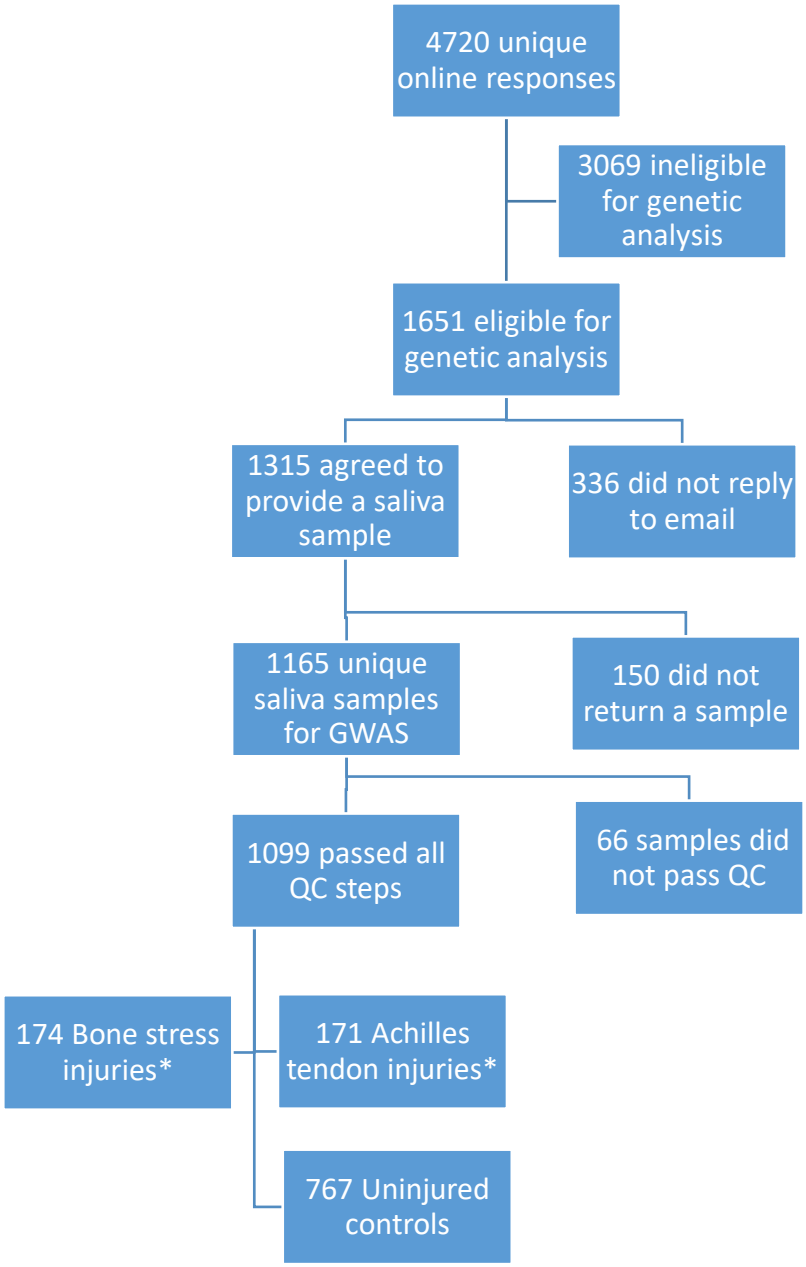


Figure 4.1 A flow chart of sample collection and subsequent filtering for case-control analysis.
* - a group comprised 13 samples with both types of injuries.

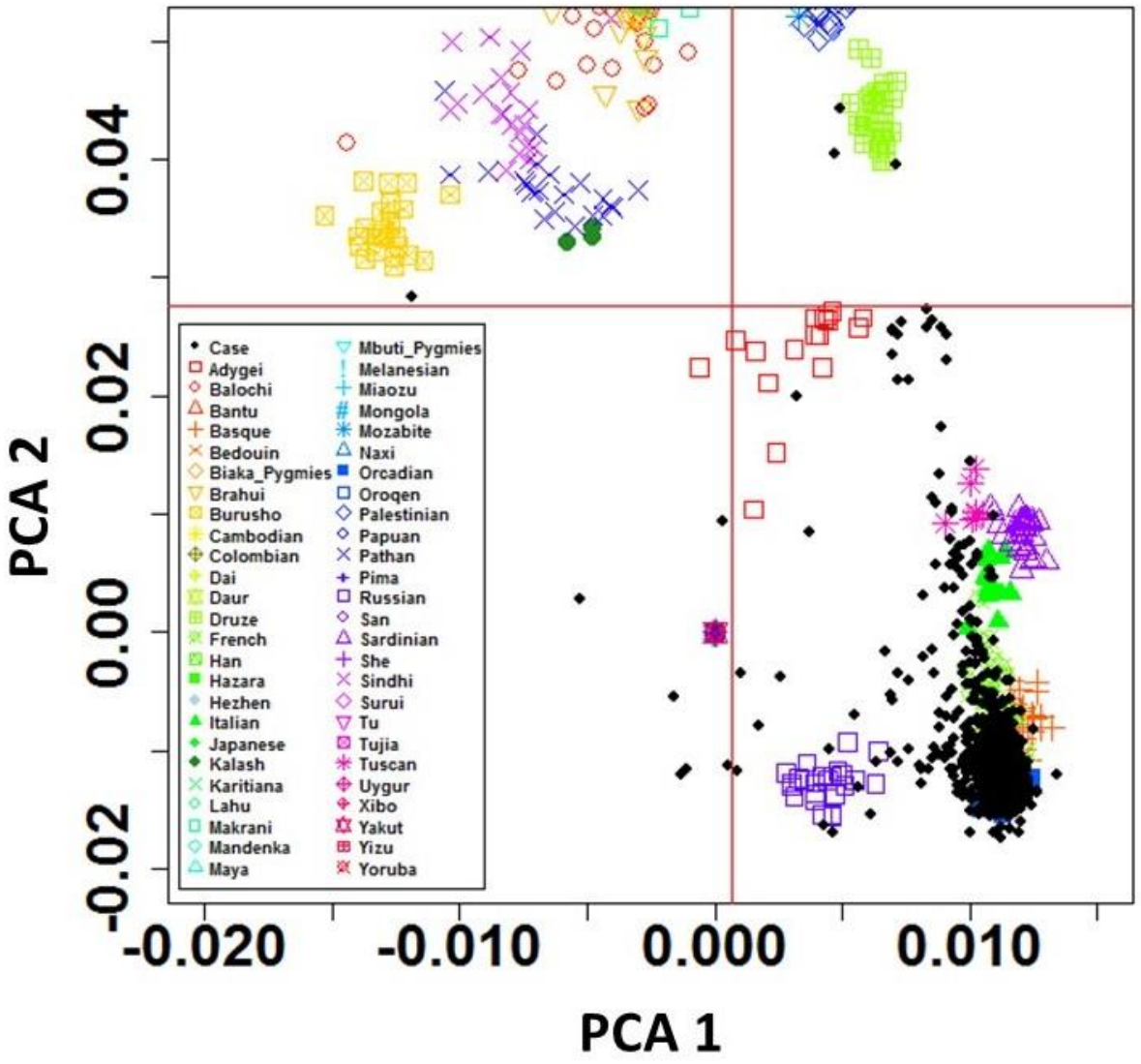


Figure 4.2 Population stratification plot.
PCA1 on the x-axis and PCA2 on the y-axis refer to the most variable components, which are commonly used to identify outliers in the population stratification analysis. Black dots identify the recreational runners used in this study, colour-coded shapes in the legend identify various ethnic groups. Red lines identify thresholds which were calculated by adding or subtracting of 6 standard deviations from the mean. The majority of the samples were located in the area of European ethnic groups and within defined thresholds. 10 samples were outliers and therefore excluded from the subsequent case-control analysis.

4.3.3 Power calculations for the collected samples

As presented in Section 4.3.1 the final numbers of participants included in the GWAS case-control analyses were: 171 AT cases, 174 BS cases and 767 uninjured controls. These participant numbers were used to calculate the actual study power compared to our targeted recruitment and study design presented in Section 1.8.3. Using a disease prevalence level of 10 , the expected significance threshold was recalculated as only 281,000 SNPs remained after the genetic data quality control and the new employed threshold was determined to be $p<2x10^{-7}$. These calculations demonstrated that with the study's sample size, SNP variants with a MAF of between 0.3 and 0.5 could be investigated if their RR was estimated within a range 1.8-2.0 (Figure 4.12). Less common SNPs with MAFs 0.2 and 0.1 could reach 80 power only if their RR were as high as 2.2 and 2.7, respectively.

These calculations demonstrated that the collected sample size was not sufficient enough to propose strong associations between identified SNPs and AT or BS injuries and discussion of any putative SNPs is predominantly explorative in nature. Thus, a MAF threshold of 0.1 was selected to explore genotyped SNPs.

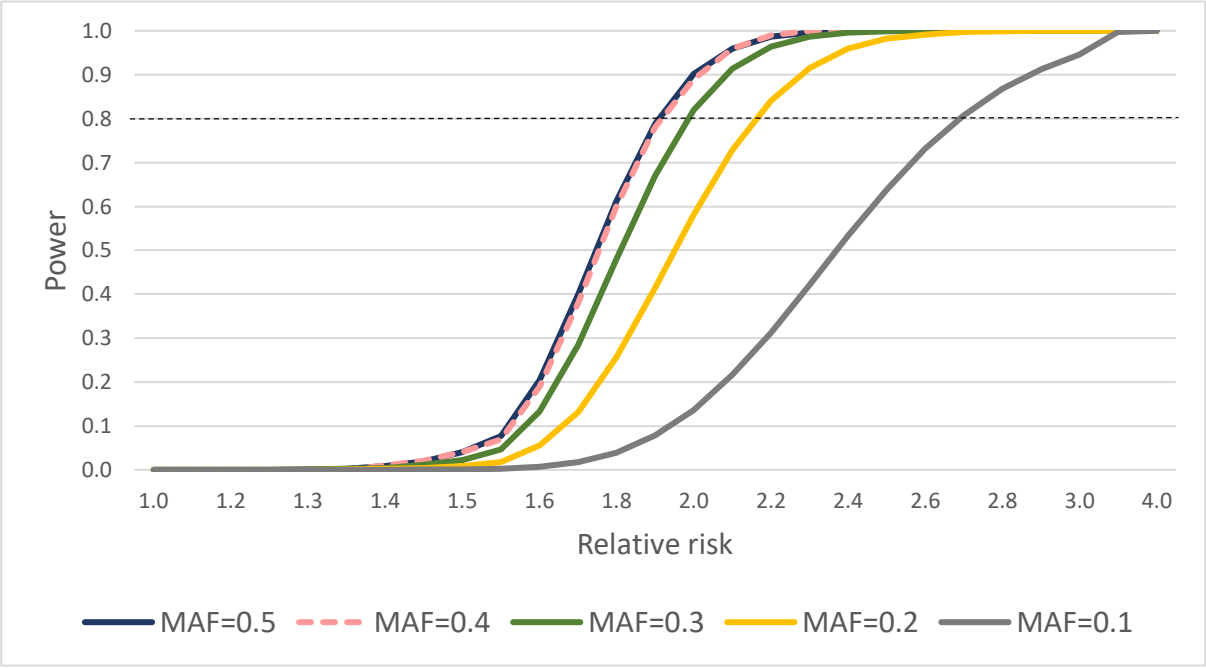


Figure 4.3 Expected relative risks of the genetic variants for 170 cases and 770 controls.

Determined values: p $2x10^{-7}$, 10 disease prevalence, MAFs varying between 0.1 and 0.5.

4.3.4 Genetic polymorphisms associated with Achilles tendon injuries

Logistic regression was used to calculate *p*-values for case-control analysis of AT genomic cohort. Following data QC, the *p*-values for 281,168 genetic variants were included in the analysis. A QQ plot of observed and theoretical *p*-values demonstrated that these values mainly overlapped, and none of the genotyped SNPs reached statistical significance ($p<5x10^{-8}$) (Figure 4.4). A Manhattan plot (Figure 4.5) shows the distribution of estimated *p*-values across 22 chromosomes and identified only one SNP close to suggesting a significance level of $p<1x10^{-5}$ on chromosome 1. It was therefore decided to investigate the top 20 most significant SNPs, and a lower threshold of $p<1x10^{-3.9}$ was employed. However, this significant decrease in *p*-value threshold dramatically increases the chance of type I errors. It was estimated that at this decreased *p*-value, 35 out of 281,168 SNPs may be false-positive associations. However, the subsequent imputation of SNPs and visualisation of the top-20 SNPs in Locus oom software allowed us to investigate each SNP and discuss the likelihood of detecting potential false-positive signals.

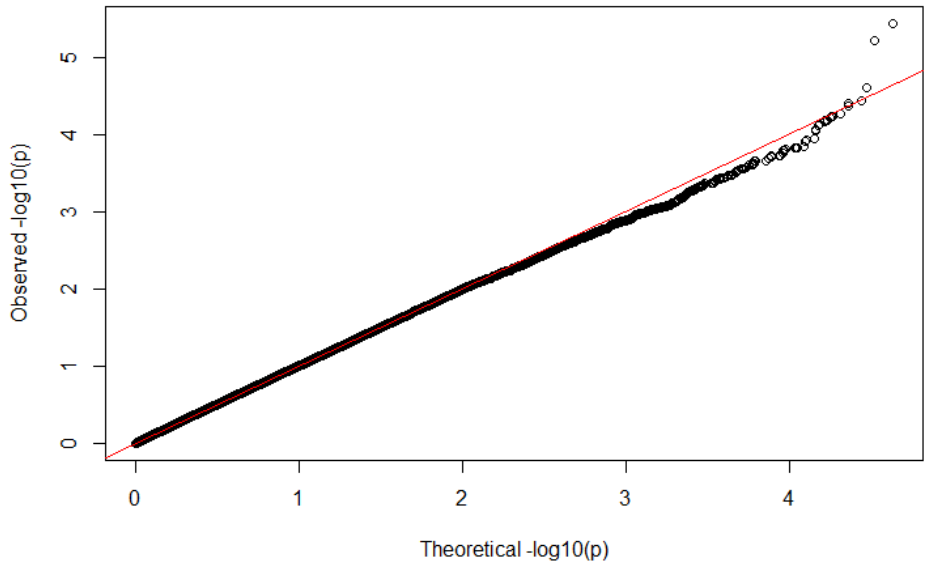


Figure 4.4 QQ plot for the Achilles tendon injury of observed and expected *p*-values, -log-transformed.
The plot shows observed *p*-values plotted on the y-axis (black dots) and theoretical *p*-values plotted on the x-axis (red line). The black dots align closely with the red line indicating that there is little to no statistical difference between the AT case and control groups.

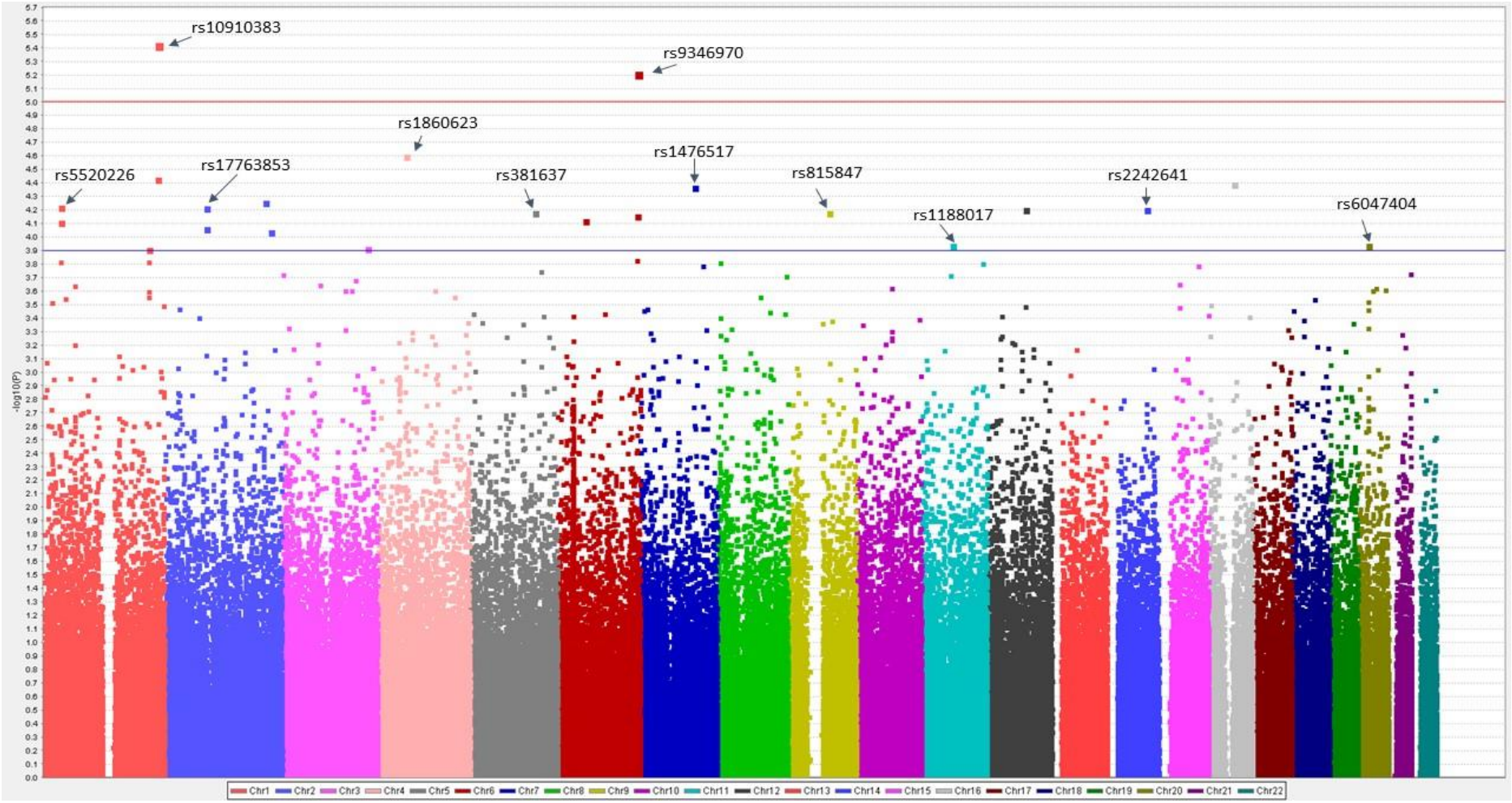


Figure 4.5 Manhattan plot of p-values calculated for the Achilles tendon injury case-control analysis, -log-transformed.
Y-axis shows –log10 *p*-value for association with an Achilles tendon injury. The red line indicates the suggestive significance threshold of $p<1x10^{-5}$, the blue line indicates the threshold of $p<1x10^{-3.9}$, which outlines the top 20 significant SNPs. Chromosome colour-coded legend is located along the x-axis. SNPs that were in the top 20 by their significance levels and located in genes, not in intragenic or uncharacterised regions are indicated on the plot.

The top 20 most significant SNPs are described in Table 4.4. The majority of these variants are located in genes; however, four SNPs were located in intragenic regions and were not in close proximity to any genes. In addition, there were four pairs of SNPs in high LD (r^2 0.6), two of these pairs were located in separate regions of chromosome 1, the third pair on chromosome 2, and the fourth pair on chromosome 6. The remaining eight SNPs appeared as single signals across eight different chromosomes. The most statistically significant SNP was rs10910383 and is located in the solute carrier family 35, member F3 (*SLC35F3*) gene (p 3.68×10^{-6}, OR 2.17) on chromosome 1. This was supported by another SNP (rs4333882) located in the same gene (p 3.69×10^{-5}, OR 1.83). Another pair of SNPs (rs6663957 and rs520226) were located in the CUB and Sushi multiple domains 2 (*CSMD2*) gene, also on chromosome 1. Chromosome 2 contained four variants, two of which (rs17763853, rs6745529) were located close to each other in the transcription factor 7 like 1 (*TCF7L1*) gene. The last pair of observed SNPs (rs9346970, rs750811) are located in the uncharacterized LOC107986666 gene on chromosome 6.

Table 4.4 Top-20 most significant genotyped SNPs, ordered by chromosomes and base-pair location.

SNP	Chr	BP	MAF	Gene	A1	OR	SE	*p*-value
rs6663957	1	34437596	0.17	*CSMD2*	A	2.95	0.27	7.63E-05
rs520226	1	34456109	0.26	*CSMD2*	A	2.94	0.27	5.89E-05
rs10910383	1	234352349	0.33	*SLC35F3*	A	2.17	0.17	3.68E-06
rs4333882	1	234352899	0.46	*SLC35F3*	G	1.83	0.15	3.69E-05
rs17763853	2	85480715	0.25	*TCF7L1*	G	0.58	0.13	5.95E-05
rs6745529	2	85503139	0.25	*TCF7L1*	A	0.61	0.12	8.52E-05
rs16839837	2	208371011	0.11	*Intragenic*	A	2.27	0.2	5.40E-05
rs3815849	2	218713469	0.01	*TNS1*	A	4.86	0.4	9.03E-05
rs1860623	4	58081590	0.37	*Intragenic*	A	1.67	0.12	2.46E-05
rs381637	5	134124658	0.43	*DDX46*	A	1.69	0.13	6.52E-05
rs61753604	6	57246878	0.03	*PRIM2*	G	2.89	0.27	7.42E-05
rs9346970	6	164109051	0.23	*LOC107986666*	C	0.37	0.22	6.00E-06
rs750811	6	164111361	0.42	*LOC107986666*	A	0.5	0.18	6.82E-05
rs1476517	7	111497826	0.29	*DOCK4*	A	1.7	0.13	4.23E-05
rs815847	9	84222618	0.5	*TLE1*	G	1.65	0.13	6.47E-05
rs1188017	11	63824364	0.09	*MACROD1*	A	1.77	0.15	1.13E-04
rs1796002	12	78923183	0.42	*Intragenic*	A	1.74	0.14	6.18E-05
rs2242641	14	80286244	0.22	*NRXN3*	A	1.65	0.13	6.16E-05
rs1110495	16	51914974	0.24	*Intragenic*	A	1.75	0.14	3.97E-05
rs6047404	20	21352236	0.28	*XRN2*	A	0.57	0.15	1.13E-04

4.3.5 Imputation of additional SNPs and visualisation of GWAS results

Quality control of imputed genotypes resulted in 23,669,623 genetic variants available for case-control and replication analyses. Case-control analysis criteria for imputed data were the same as for genotyped data, resulting in the calculation of *p*-values for 7,438,753 genetic variants. The imputed data was used to generate more detailed plots around the previously identified 20 most significant SNPs (Table 4.4). As shown in Table 4.4, three genes were supported by two SNPs each, specifically *SLC35F3* (rs10910383 and rs4333882), *CSMD2* (rs6663957 and rs520226) and *TCF7L1* (rs6745529 and rs17763853). Imputation of SNPs around these loci showed they were in strong LD with other SNPs located in the same genes (Figure 4.6A-C). Thus, despite the determined low *p*-value threshold, these results were less likely to be false-positive discoveries and were included in the following discussion. The rs9346970 and rs750811 SNPs (*LOC107986666*) were also shown to be in high LD with other local SNPs. This gene encodes for a non-coding RNA of uncharacterised function. It is therefore difficult to explain any possible link between this genomic region and susceptibility to Achilles tendon injuries (Figure 4.6D).

Locus oom plots of the SNPs that appeared as single signals on chromosomes 7, 9, 11 and 14 (rs1476517, rs815847, rs1188017, rs2242641, respectively) showed consistent LD patterns of the supporting SNPs, located in the same genes as genotyped SNPs (Figure 4.7A-D). Interestingly, rs1188017 was located in a region with a high density of genes, and supporting SNPs were also located in MACRO domain containing 1 (*MACROD1*) gene. The MAF of rs1188017 was 0.9, which was lower than defined threshold of 10 , however, it was decided to include the *MACROD1* gene in the following discussion, as MAF was close to 10 , and the function of this gene, as it is involved in regulation of transcriptional activity of nuclear factor κB (NF-κB) (Wu et al., 2011). Therefore, four genes: *MACROD1*, dedicator of cytokinesis 4 (*DOCK4*), transducing-like enhancer of split 1 (*TLE1*) and neurexin 3 (*NRXN3*) were also considered as potentially important genes in the development of Achilles tendon injuries.

Plots of two SNPs (rs3815849, rs61753604) showed no LD patterns, and the plots of the other two SNPs (rs381637, rs6047404) had LD patterns scattered across surrounding genetic regions (Figure 4.8A-D). The former two SNPs (rs3815849, rs61753604) are located in tensin 1 (*TNS1*) and DNA primase subunit 2 (*PRIM2*) genes, respectively. As both of these SNPs also had low MAF (<10), these two variants were excluded from the

subsequent investigation. The latter two SNPs (rs381637, rs6047404) were located in the DEAD-box helicase 46 (*DDX46*) and 5'-3' exoribonuclease 2 (*XRN2*) genes, respectively. These two genes were included in the subsequent discussion of identified SNPs.

The last four plotted SNPs were located in intragenic regions, and Locus oom plots did not show any SNPs in high LD, which would be located in genes in the surrounding regions (Figure 4.9A-D). All these SNPs were excluded for the subsequent discussion.

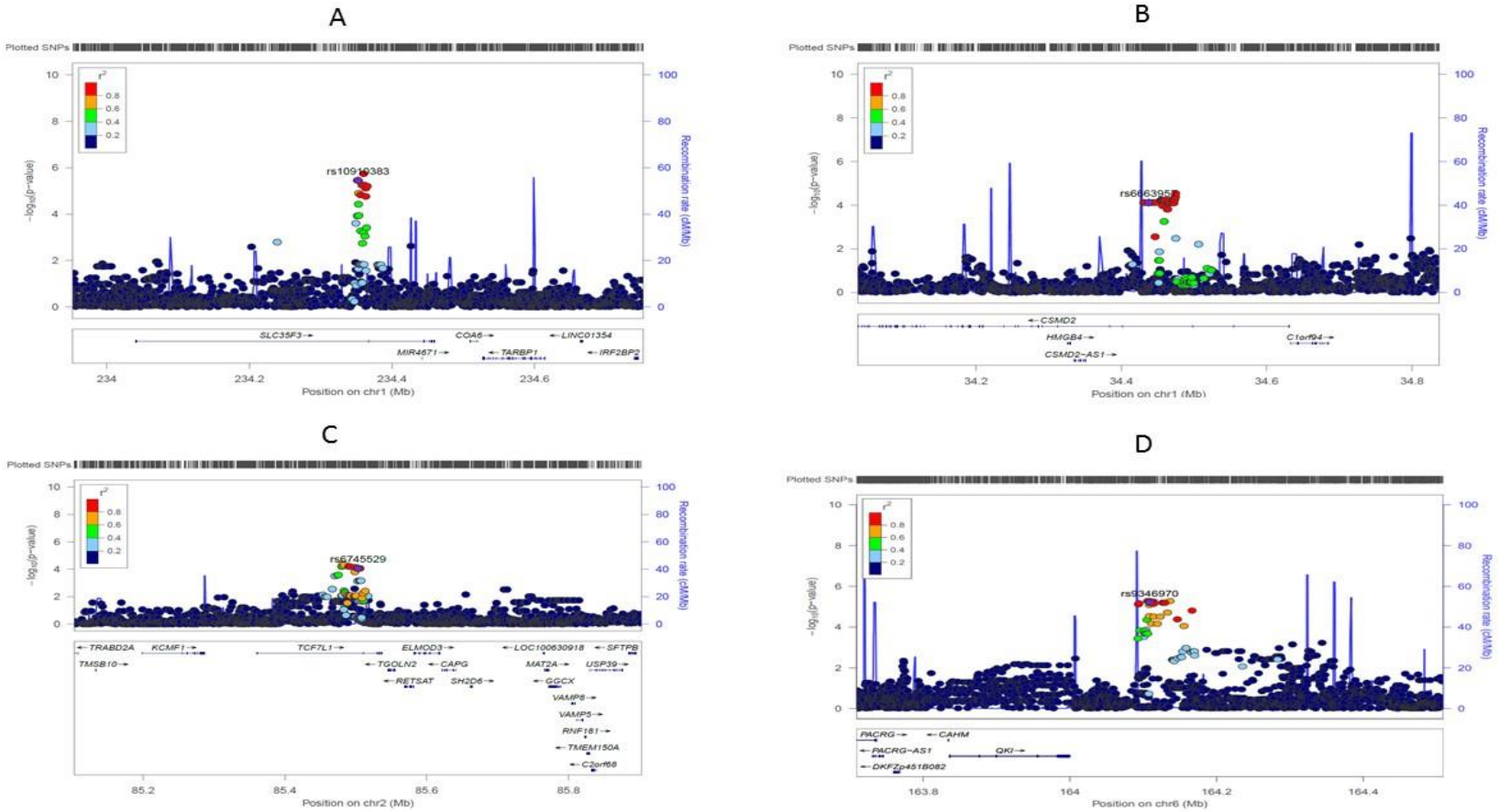


Figure 4.6 Locus Zoom plots for identified significant SNPs from the paired signals.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 1 in *SLC35F3* gene; B) signals on chromosome 1 in *CSMD2* gene; C) signals on chromosome 2 in *TCF7L1* gene D) signals on chromosome 6 in *LOC107986666.*

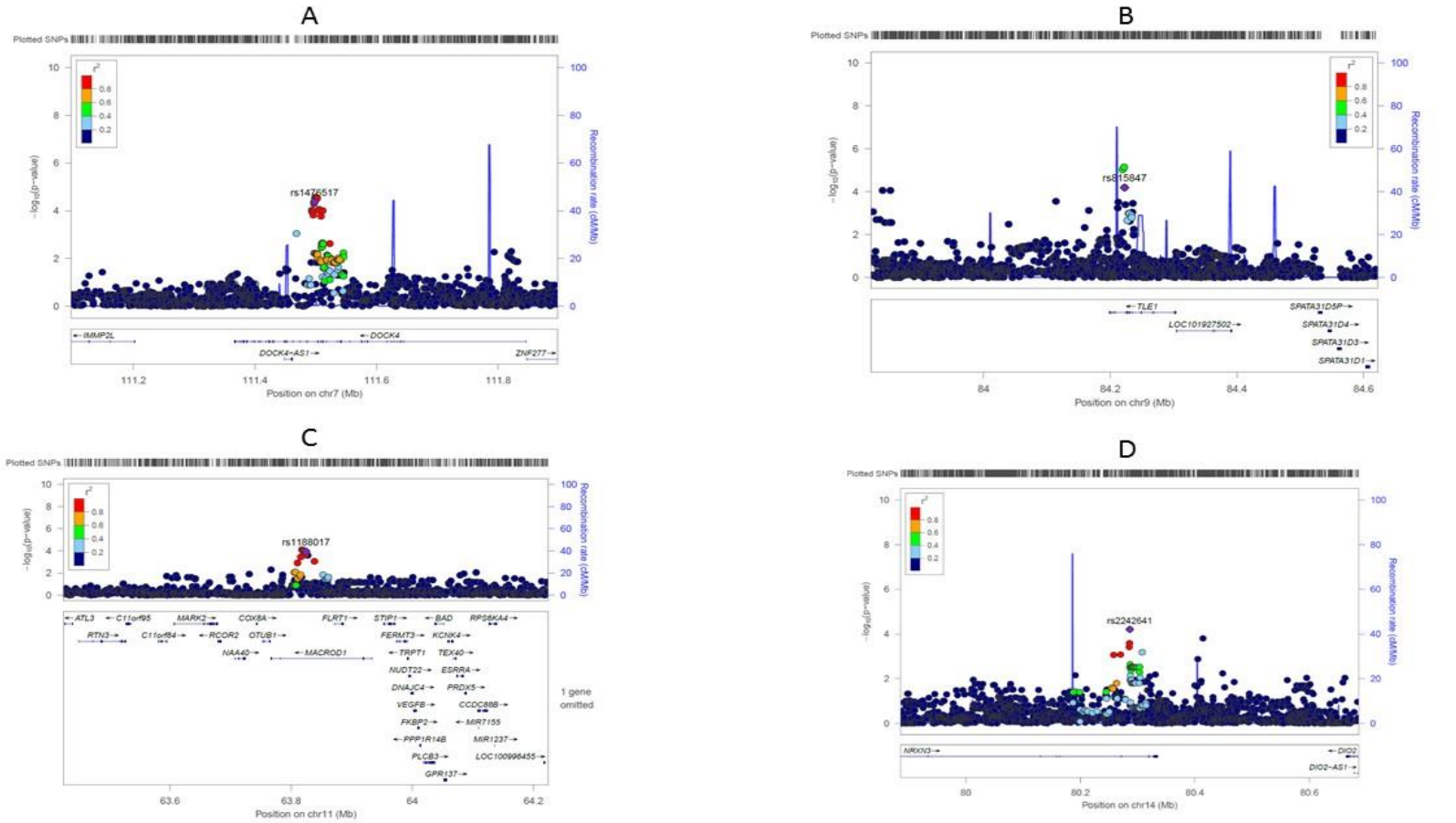


Figure 4.7 Locus Zoom plots for identified significant SNPs from chromosomes 7, 9, 11 and 14.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 7 in *DOCK4* gene; B) signals on chromosome 9 in *TLE1* gene; C) signals on chromosome 11 in *MACROD1* gene; D) signals on chromosome 14 in *NRXN3* gene.

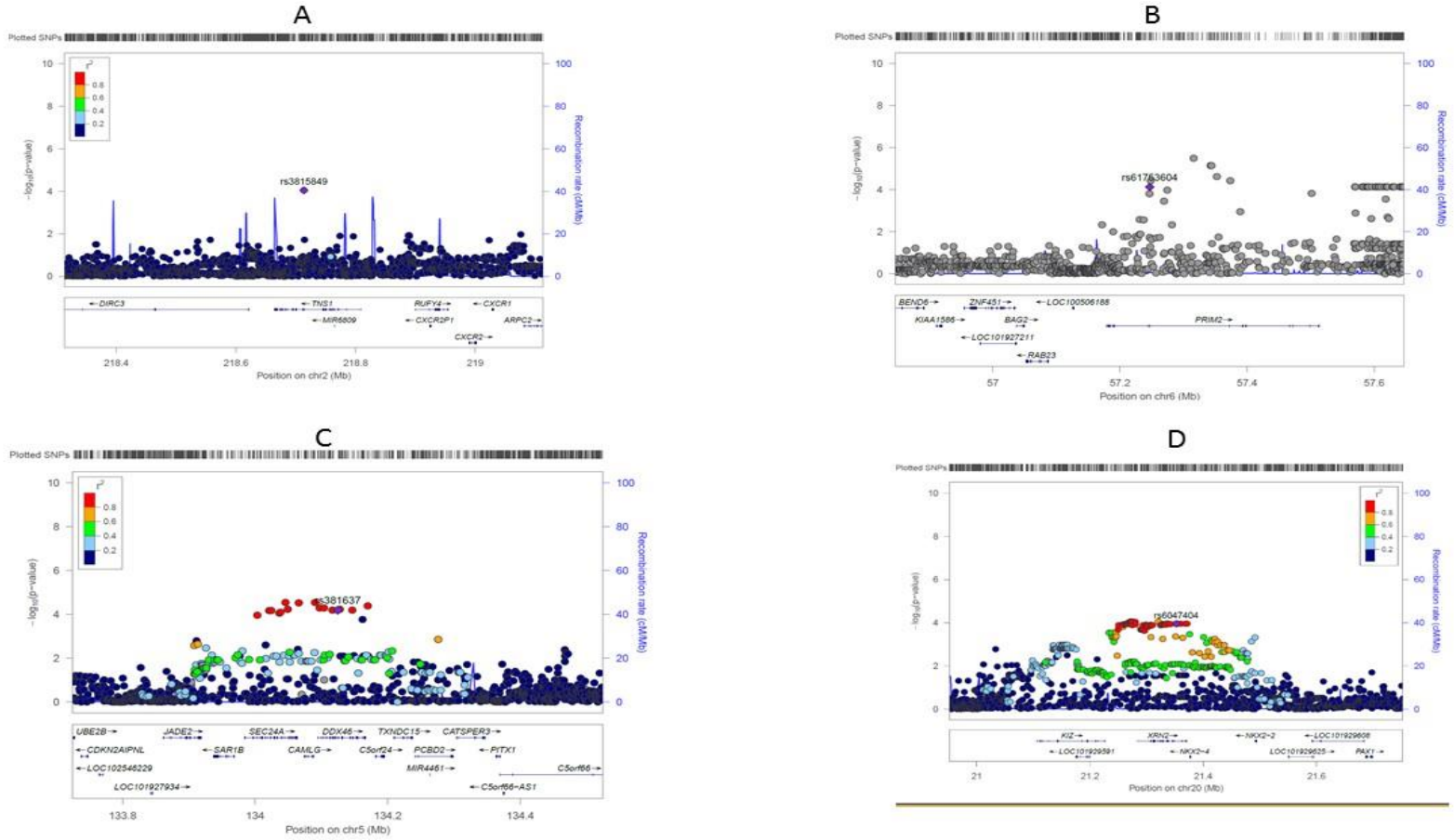


Figure 4.8 Locus Zoom plots for identified significant SNPs from chromosomes 2, 6, 5 and 20.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 2 in *TNS1* gene; B) signals on chromosome 6 in *PRIM2* gene; C) signals on chromosome 5 in *DDX46* gene; D) signals on chromosome 20 in *XRN2* gene.

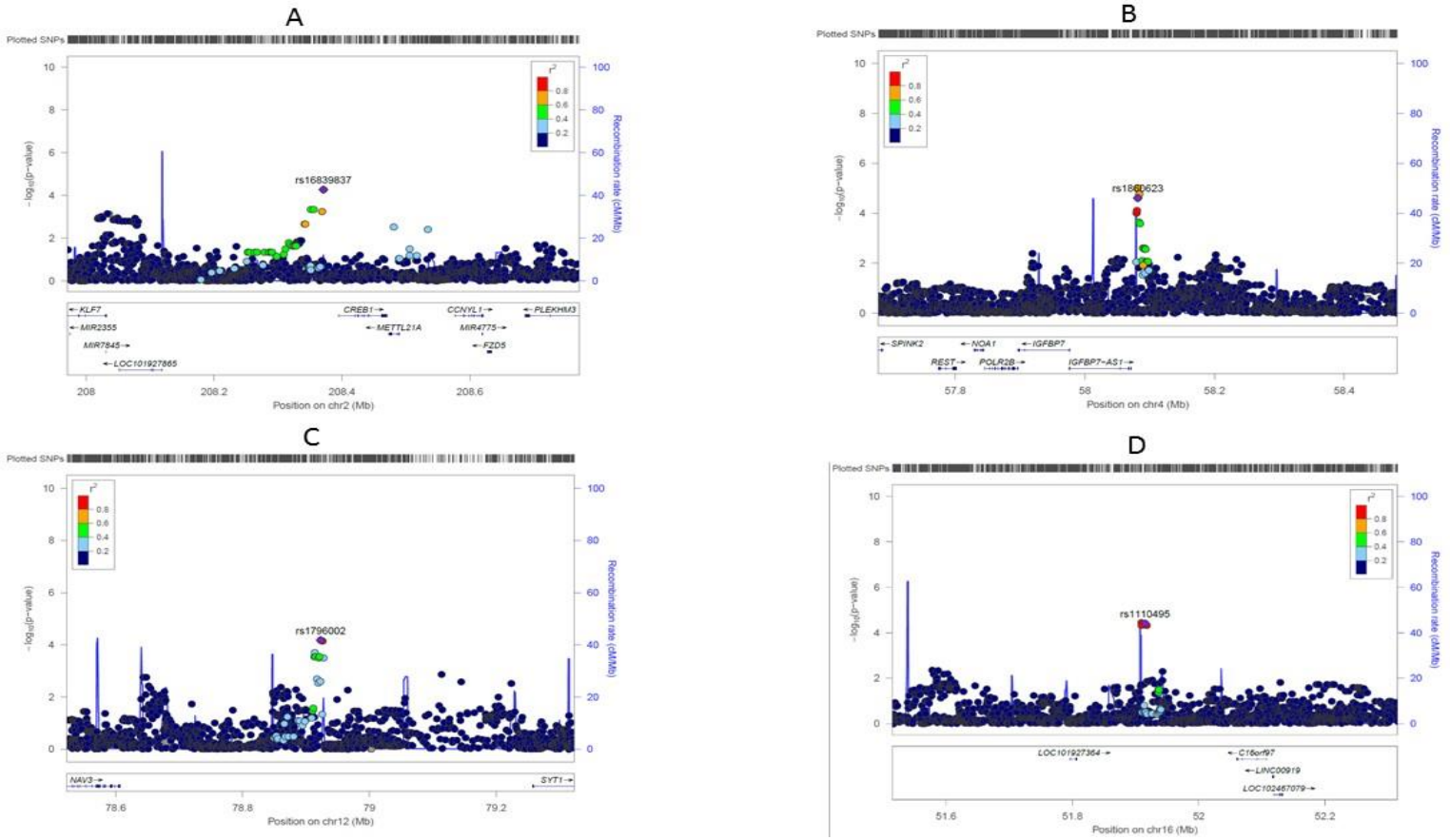


Figure 4.9 Locus Zoom plots for identified significant SNPs from chromosomes 2, 4, 12 and 16.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 2 in the intragenic region; B) signals on chromosome 4 in the intragenic region; C) signals on chromosome 12 in the intragenic region; D) signals on chromosome 16 in the intragenic region.

4.3.6 Replication analysis

Imputation of additional SNPs allowed us to attempt replication analysis of the top four significant SNPs previously identified in the GWAS by Kim et al.; identify frequencies of genotyped top 20 SNPs in the publicly available data from Kim et al. GWAS; identify *p*-values of SNPs investigated in the previously described studies, which used a candidate gene approach.

4.3.6.1 Replication analysis of top four significant SNPs using data from GWAS by Kim et al.

None of the four SNPs (rs57104447, rs60713544, rs1937810, rs57224706) that reached suggestive significance levels in the GWAS by Kim et al., were identified in our cohort using the imputed dataset.

4.3.6.2 Replication of top-20 SNPs using data from GWAS by Kim et al.

The attempted replication analysis of the 20 most significant SNPs was performed using the publicly available Kim et al. data (National Heart, 2017). However, data on two SNPs (rs61753604, rs1796002) were not available in the accessed data. As shown in Table 4.5, none of the top 20 most significant SNPs identified in our cohort were at the suggested significance level in the Kim et al. study cohort.

Table 4.5 Summary of *p*-values identified for top 20 genotyped SNPs in the dataset from Kim et al.

SNP	Chr	Gene	OR	*p*-value
rs6663957	1	*CSMD2*	1.04	0.4412
rs520226	1	*CSMD2*	1.04	0.3765
rs10910383	1	*SLC35F3*	0.1	0.7260
rs4333882	1	*SLC35F3*	0.1	0.7073
rs17763853	2	*TCF7L1*	1.0	0.8652
rs6745529	2	*TCF7L1*	1.0	0.9927
rs16839837	2	Intragenic	1.0	0.1772
rs3815849	2	*TNS1*	0.92	0.4624
rs1860623	4	Intragenic	0.98	0.4086
rs381637	5	*DDX46*	0.97	0.8092
rs61753604	6	*PRIM2*	*n.d.*	*n.d.*
rs9346970	6	*LOC107986666*	1.03	0.3618
rs750811	6	*LOC107986666*	1.03	0.2029
rs1476517	7	*DOCK4*	1.02	0.4466
rs815847	9	*TLE1*	0.98	0.2549
rs1188017	11	*MACROD1*	0.95	0.3574
rs1796002	12	Intragenic	*n.d.*	*n.d.*
rs2242641	14	*NRXN3*	1.02	0.4091
rs1110495	16	*N/A*	0.95	0.0215
rs6047404	20	*XRN2*	1.01	0.8361

n.d. not determined

4.3.6.3 Replication analysis of SNPs identified in the studies using a candidate gene approach

Previous studies investigating candidate genes have found several SNPs significantly associated with Achilles tendon injuries. These SNPs were included in replication analysis and imputed data were used to identify all *p*-values. A summary of *p*-values for 33 SNPs located in 25 genes investigated in relation to Achilles tendon injuries is presented in Table 4.6. None of these 33 SNPs were associated with Achilles tendon injuries in our cohort, hence there was no strong evidence of replication of previously reported significant associations. While the rs12722 variant in the *COL5A1* gene did reach putative significance (*p* 0.0194), this result should be interpreted with caution due to its low *p*-value and risk of type I error. Caution should also be taken as these data have been obtained through imputation, and require further confirmation by direct genotyping for more conclusive evidence of association. Locus oom plots for rs12722 (*COL5A1*), rs2104772 (*TNC*), rs1045485 (*CASP8)*, rs650108 (*MMP3*), rs4789932 (*TIMP2*), rs143383 (*GDF5*) are displayed in the Figures 4.10 A-D and 4.11 A-B, respectively. The statistical significance levels reached by these SNPs were low. Additionally, the significance levels of SNPs shown to be in LD pattern with replicated SNPs were also low. These plots confirmed the absence of any significant SNPs in the genetic regions containing genes reported to be significant by studies that used a candidate gene approach. This outcome, therefore, confirms the results of Kim et al. GWAS, which also was unable to replicate the association of these SNPs (Kim et al., 2017).

Table 4.6 Summary of candidate SNPs and their significance levels identified using imputed data.

SNP	Chr	BP	MAF	Gene	A1	OR	SE	*p*-value
rs3753841	1	103379918	0.5	*COL11A1*	G	0.88	0.13	0.3176
rs1676486	1	103354138	0.22	*COL11A1*	A	0.99	0.15	0.9278
rs1045485	2	202149589	0.05	*CASP8*	C	0.85	0.18	0.3485
rs1143627	2	113594387	0.47	*IL1B*	G	1.07	0.13	0.6006
rs1049253	4	185548951	0.08	*CASP3*	G	0.99	0.15	0.9261
rs1054480	5	178540975	0.26	*ADAMTS2*	A	1.0	0.13	0.9816
rs331079	5	127770805	0.11	*FBN2*	C	0.89	0.2	0.5584
rs1799907	6	33152835	0.32	*COL11A2*	T	1.11	0.13	0.4123
rs240736	6	75848181	0.27	*COL12A1*	G	1.09	0.13	0.5080
rs970547	6	75797302	0.29	*COL12A1*	C	1.11	0.15	0.4823
rs2071307	7	73470714	0.22	*ELN*	A	1.05	0.12	0.7193
rs1800795	7	22766645	0.14	*IL-6*	C	1.02	0.12	0.8493
rs1799983	7	150696111	0.18	*NOS3*	T	1.1	0.13	0.4692
rs4870723	8	121228679	0.41	*COL14A1*	C	0.89	0.12	0.3229
rs1563392	8	121353692	0.48	*COL14A1*	T	1.0	0.12	0.9702
rs946053	9	117049891	0.24	*COL27A1*	T	0.97	0.12	0.7891
rs12722	9	137734416	0.35	COL5A1	C	1.32	0.12	0.0194
rs1134170	9	137735274	0.39	COL5A1	A	1.15	0.18	0.2768
rs13321	9	117792583	0.33	*TNC*	C	0.98	0.13	0.8503
rs2104772	9	117808785	0.48	*TNC*	A	1.16	0.12	0.2249
rs1330363	9	117813990	0.42	*TNC*	C	0.99	0.12	0.9057
rs3740199	10	128019025	0.48	*ADAM12*	C	1.09	0.12	0.5000
rs679620	11	102713620	0.35	*MMP3*	C	1.04	0.12	0.7580
rs591058	11	102711338	0.36	*MMP3*	C	1.05	0.12	0.6840
rs650108	11	102708787	0.44	*MMP3*	A	0.88	0.14	0.3578
rs4149577	12	6447522	0.4	*TNFRSF1A*	G	0.97	0.12	0.7865
rs2779249	17	26128581	0.27	*NOS2*	A	1.0	0.13	0.9922
rs4789932	17	76924275	0.44	*TIMP2*	A	1.13	0.12	0.3194
rs1800469	19	41860296	0.37	*TGFB1*	A	1.03	0.13	0.8189
rs143383	20	34025983	0.45	*GDF5*	G	1.13	0.13	0.3389
rs226794	21	28302355	0.19	*ADAMTS5*	A	0.83	0.2	0.3599

SNP – single nucleotide polymorphism, Chr – chromosome, BP – base pairs, MAF – minor allele frequency, A1 – effect allele, OR – odds ratio, SE – standard error.

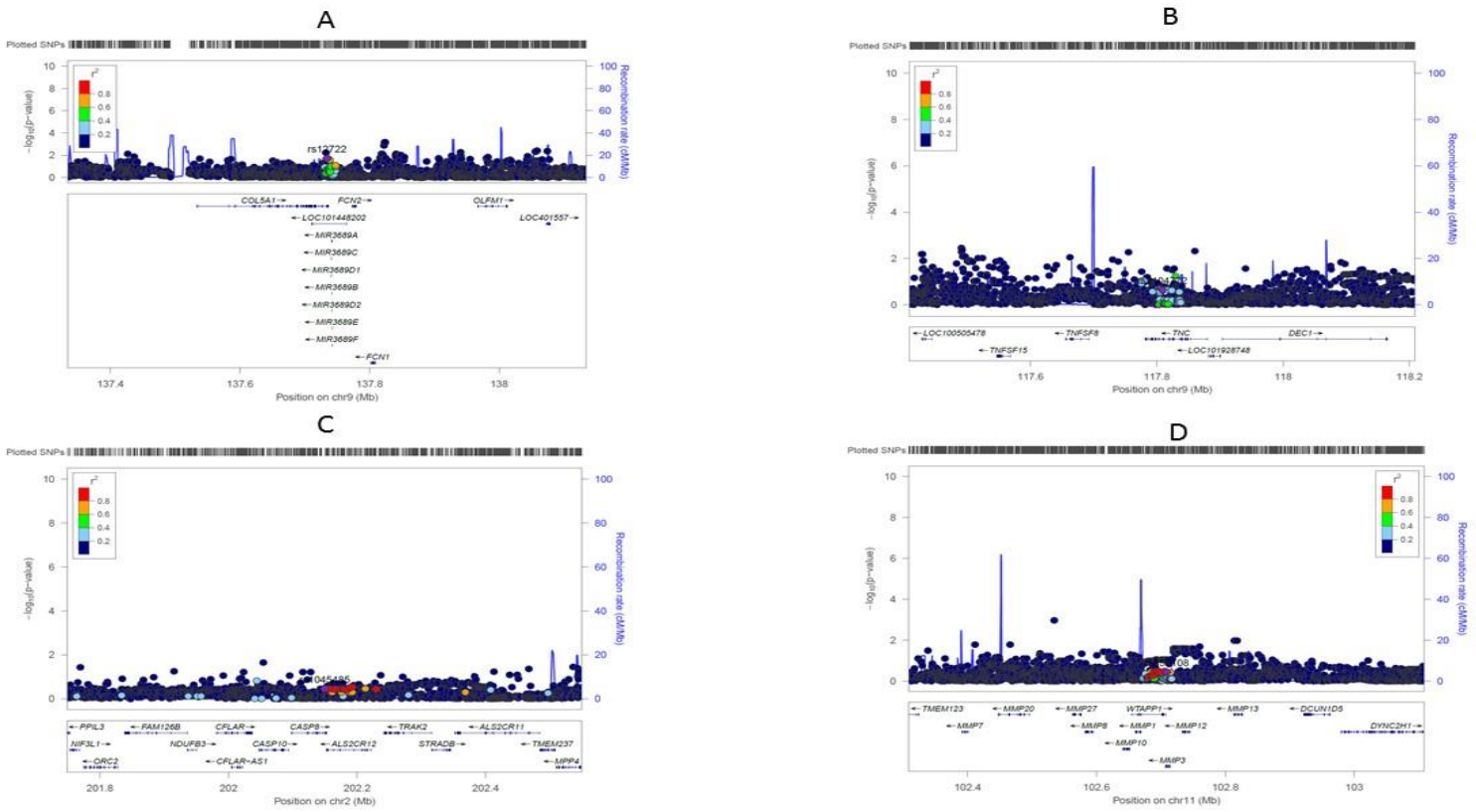


Figure 4.10 Locus Zoom plots for imputed SNPs included in replication analysis from chromosomes 9, 2 and 11.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 9 in *COL5A1* gene; B) signals on chromosome 9 in *TNC* gene; C) signals on chromosome 2 in *CASP8* gene; D) signals on chromosome 11 in *MMP3* gene.

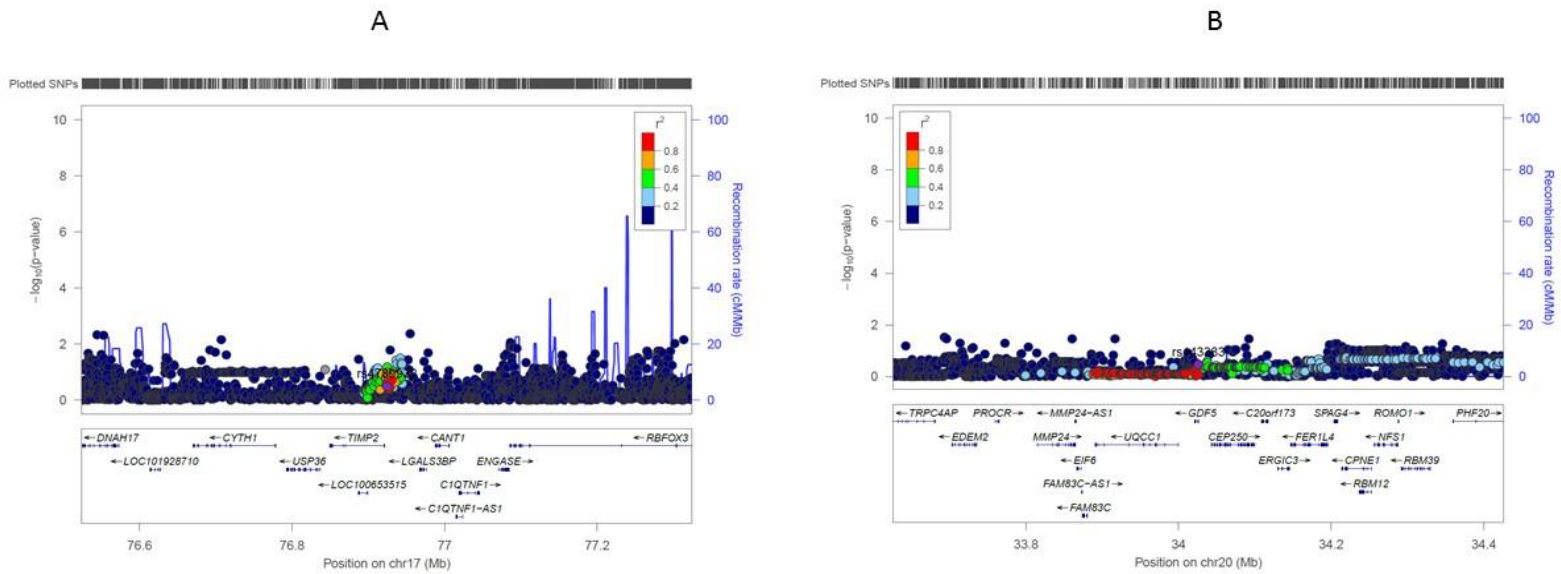


Figure 4.11 Locus Zoom plots for imputed SNPs included in replication analysis from chromosomes 17 and 20.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 17 in *TIMP2* gene; B) signals on chromosome 20 in *GDF5* gene.

4.4 Discussion

4.4.1 Genetic case-control analysis

Achilles tendinopathy is a multifactorial condition and one of the most common running-related injuries (Dias Lopes et al., 2012). Although training load plays a major role in the development of the pathology, it is modulated by other intrinsic factors such as sex, age, anatomy and genetic predisposition (Cook & Purdam, 2009). Previously investigated genetic polymorphisms located in genes encoding certain types of collagen and proteins involved in extracellular matrix metabolism showed significant association with the development of Achilles tendinopathy (Collins, 2010). However, these small numbers of genetic polymorphisms were analysed in small, limited cohorts with an average of 150 cases. Thus, development of large-scale approaches for genetic analysis, such as GWAS allowed screening of the whole genome and a search for genetic variants associated with multifactorial conditions such as Achilles tendinopathy. This GWA approach is unbiased and is more likely to identify the strongest associations between genetic polymorphisms and injury risk than previously utilised the candidate gene approach.

The current GWAS was conducted on 938 samples, comprising of 171 Achilles tendon injury cases and 767 uninjured controls. All included individuals were physically active, running at least 15 km per week. However, case-control analysis of the genetic data did not identify any statistically significant polymorphisms at the expected for GWAS of $p<5x10^{-8}$, this was not unexpected, given the lower than expected recruitment and power analysis. Therefore, it was decided to explore the 20 most significant genetic variants, which corresponded to above $p<5x10^{-3.9}$ threshold. Whilst this exploratory approach increases the likelihood of type I errors, the additional imputation of SNPs and visualisation of the results allowed us to critically assess each signal and identify the most probable SNPs for investigation. Furthermore, the corresponding protein function was also considered to support the putative association with Achilles tendon injury development. Among these 20 SNPs, there were four paired SNPs located in *CSMD2*, *SLC35F3*, *TCF7L1* and *LOC107986666*. After plotting genetic regions surrounding these genes, the function and the potential role of these genes in the development of the Achilles tendon injury were further explored. After revision of the remaining single SNPs and their plots, a further six genes were also considered as

potential contributors to the development of the Achilles tendon injury, specifically *DOCK4, DDX46, TLE1, MACROD1, NRXN3* and *XRN2*.

Transcription factor 7 like 1 (*TCF7L1*) is a gene the product of which is a member of T cell factor/lymphoid enhancer factor (Tcf/lef) family, activated by β-catenin and thus mediates the Wnt signalling pathway (Figure 4.12). This pathway is highly conserved and targets multiple genes, including MMPs (Clevers, 2006). MMPs regulate extracellular matrix homeostasis and have been previously implicated in Achilles tendinopathy, with *MMP3* polymorphisms previously shown to be associated with Achilles tendinopathy (Raleigh, 2009). Bioinformatic analysis of an upstream region of the *MMP3* gene showed a presence of a Tcf/lef binding site, also typical for many other MMPs (Clark, Swingler, Sampieri, & Edwards, 2008). This finding requires further research into whether a genetic variation of the *TCF7L1* gene may affect the regulation of expression of *MMP3* through the Tcf/lef transcription complex. Additionally, the Wnt pathway regulates the development of cartilage and skeletal system. Genetic studies showed that alterations in the expression of molecules involved in the Wnt/β-catenin pathway were associated with osteoarthritis and therefore may be a therapeutic target (Usami, Gunawardena, Iwamoto, & Enomoto-Iwamoto, 2016).

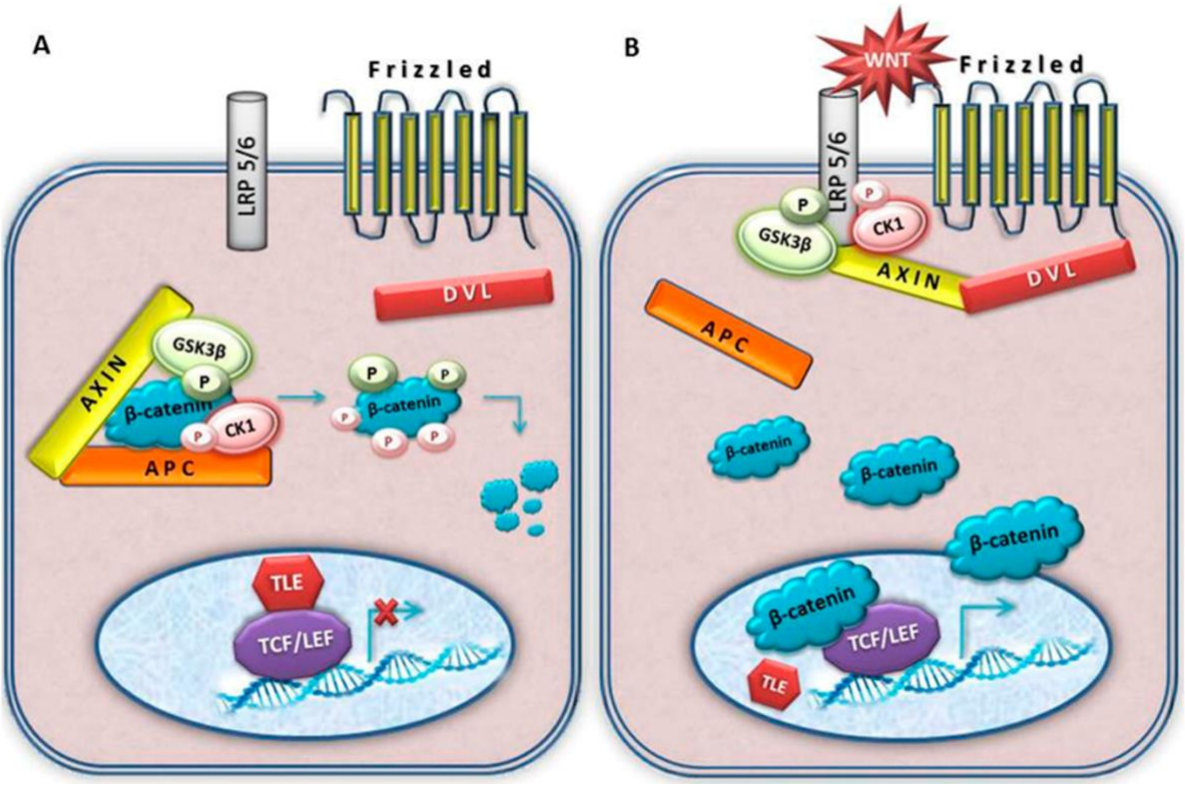


Figure 4.12 A schematic illustration of the canonical Wnt pathway.
Panel (A), in the absence of Wnt ligand, a destruction complex consisting of AXIN, APC, GSK3-β and CK1 resides in the cytosol. β-catenin is phosphorylated by CK1 and GSK3-β and targeted for degradation by the proteasomal machinery; Panel (B), with Wnt stimulation, some components of protein complex dislocate from the cytosol to the plasma membrane. The destruction complex falls apart, and β-catenin is stabilised. Dvl is also recruited to the membrane and binds to Fz and Axin, which is bound to phosphorylated LRP5/6. Stabilised β-catenin is translocated to the nucleus where it associates to LEF/TCF transcription factors, displacing co-repressor TLE and recruiting additional co-activators to Wnt target genes. Adapted from Pecina-Slaus et al. (Pećina-Šlaus, Kafka, & Lechpammer, 2016).

Dedicator of cytokinesis 4 (*DOCK4*) is a gene located at chromosome 7q31.1. The encoded protein DOCK4 activates GTPase Rac1, which in turn is responsible for nuclear translocation of β-catenin and Tcf activation, and therefore regulates the Wnt signalling pathway (Wu et al., 2008). It has previously been shown that DOCK4 enhances β-catenin stability and Tcf activation, therefore regulating the Wnt/β-catenin signalling pathway (Upadhyay et al., 2008). This function of DOCK4 may potentially affect the expression of MMPs and other target genes of the Wnt pathway, which may be involved in molecular processes of extracellular matrix homeostasis in the tendon. However, polymorphisms in *DOCK4* and its expression were predominantly

investigated in relation to tumorigenesis, particularly ovarian cancer and myelodysplastic syndromes (Kuo et al., 2009; Sundaravel et al., 2015).

Transducin-like enhancer of split 1 (*TLE1*) is a gene encoding a transcriptional co-repressor and is bound to the Tcf/lef complex, which can block transcription of the Wnt target genes. When the Wnt pathway is activated, β-catenin directly displaces TLE repressor from the Tcf/lef complex and thus activates transcription of the target genes (Daniels & Weis, 2005). Since TLE1 is involved in regulation of expression of various genes, it was investigated in relation to neuron differentiation, its expression was shown in macrophages and used as a diagnostic immunochemical marker for synovial sarcomas (Bakrin, Hussain, & Tuan, 2016; Buscarlet et al., 2009; De Paoli et al., 2016). However, *TLE1* has not been studied in association with connective tissue.

CUB and Sushi multiple domains 2 (*CSMD2*) is a gene encoding protein with a Sushi multiple domain feature, which characterises it as a regulator of the complement cascade – a signalling pathway of the innate immune system. This protein, therefore, helps to clear pathogens or tag them for subsequent destruction by other cells. Polymorphisms in this gene were investigated in relation to schizophrenia, as this gene is expressed in brain tissue. A study identified eight SNPs located in *CSMD2*, which were in a strong association with schizophrenia (H vik et al., 2011). However, none of these eight SNPs matched either of two *CSMD2* SNPs identified in this study. In addition, expression of *CSMD2* is shown to be decreased in cancer tissue of patients with colorectal cancer, as well as expression of other two genes of CSMD family (*CSMD1 and CSMD3*) (hang & Song, 2014). This finding suggested that CSMD genes play a tumour suppressing role and may be used as predicting markers of colorectal cancer. In the current study, two polymorphisms in *CSMD2* were associated with an increased risk of Achilles tendon injury. The role of inflammation in tendon pathology is complex, and although inflammatory cells are observed in pathological tendons, as well as an increase of inflammatory cytokines, this inflammatory response is not a primary event in the development of pathology. It was shown that tenocytes express cytokines in response to cyclic load (Cook, Rio, Purdam, & Docking, 2016). Potentially, two SNPs of the *CSMD2* gene were among the top 20 signals because its product is involved in the immune response and regulation of complement cascade, which triggers expression of cytokines.

Solute carrier family 35, member F3 (*SLC35F3*) is a gene which encodes solute transporter catalyses thiamine (vitamin B-1) transport and was investigated in relation to hypertension. Two *SLC35F3* SNPs were significantly associated with hypertension in a GWAS (hang et al., 2014). However, these significant SNPs differed from those identified in the current study. Thiamine deficiency causes a spectrum of phenotypes, including cardiovascular and neurological manifestations of beriberi. In a rat model, it was shown that thiamine repletion corrected glucose oxidation, elevated blood pressure and decreased expression of angiotensin-converting enzyme (ACE), angiotensin (AGT) and angiotensin II receptor (AGTR1) (Tanaka et al., 2007). Previously it was shown that joint functioning of ACE, AGT and angiotensin II receptor induces increased synthesis of protein of extracellular matrix and increased levels of connective tissue growth factor and TGFB1 in skeletal muscles (Cabello-Verrugio et al., 2011; Morales et al., 2012). Inhibition of ACE decreased fibrosis of skeletal muscles in dystrophic mice (Morales et al., 2013). Extracellular matrix disorganisation is characteristic for pathological tendons (Cook & Purdam, 2009). Although ACE functions were not investigated in relation to tendon pathology, we could speculate that the investigated gene *SLC35F3* may distantly affect ACE functioning and therefore metabolism of the extracellular matrix.

MACRO domain containing 1 (*MACROD1*) gene is located at the chromosome 11q13.1. The encoded protein contains a highly conserved macro domain and is also known as LRP16. It was shown that the transcriptional activity of the nuclear factor κB (NF-κB) is regulated through the interaction between LRP16 and UXT (an α-class prefoldin protein) (Wu et al., 2011). NF-κB plays a role in multiple processes such as inflammation, cell survival and immunity and therefore its activity was associated with multiple human diseases that affect processes: arthritis, diabetes, cancer, inflammatory bowel disease (Baldwin Jr, 2001). Additionally, LRP16 has a specific function as a coactivator of both estrogen receptor α and androgen receptor, interestingly, at the same time, it is their target. This explains why LRP16 was extensively studied in hormone-dependent types of cancer (Wu et al., 2011). NF-κB is a mediator of cytokine expression (Kulms & Schwarz, 2006), which are elevated in the pathological tendon, but the development of inflammation is not necessarily a key driver event in tendinopathy (Cook et al., 2016). Therefore, NF-κB function and regulation may play a role in the inflammatory process, which may accompany

tendinopathy. However, the direct involvement of NF-κB has not currently been investigated in association with tendon pathology in detail.

DEAD-box helicase 46 (*DDX46*) encodes a protein which belongs to a family of RNA helicases, which contain a highly conserved helicase domain. These proteins are responsible for mRNA splicing and ribosome and spliceosome assembly (Hirabayashi, Hozumi, Higashijima, & Kikuchi, 2013). DDX46 pre-mRNA splicing is required during the development of the digestive system and the brain (Hozumi et al., 2012). DDX46 was also involved in the development of several tumours, such as oesophageal squamous cell carcinoma and osteosarcoma (Jiang, hang, Li, & Wang, 2017). However, the DDX46 gene's function has not been investigated in tendon cells or in relation to any tendon pathology.

5'-3' exoribonuclease 2 (*XRN2*) is a gene located at chromosome 20p11.22 and encodes for a 5'-3' exonuclease involved in transcription termination (Rother et al., 2016). XRN2 forms a complex with NF-κB-repressing factor (NKRF) and DEAH-box RNA helicase (DHX15) to degrade excised ribosomal pre-rRNA spacer fragments (Memet, Doebele, Sloan, & Bohnsack, 2017). Due to its function of transcription terminator, XRN2 was identified as an important factor for the maintenance of genomic stability and cell survival (Morales et al., 2016). Although XRN2 is involved in crucial for living cell processes, it has never been investigated in association with tendon pathology.

Neurexin 3 (*NRXN3*) is a protein-coding gene located on chromosome 14q24.3-q31.1. This gene is the largest of three neurexin genes with the most extensive possible alternative splicing, resulting in 1,764 transcript variants and leads to a high diversity of the proteins. These proteins function in the nervous system as receptors and cell adhesion molecules (Rowen et al., 2002). Polymorphisms in *NRXN3* have been studied extensively in association with various neurodevelopmental conditions (autism spectrum disorder), addictive behaviours and borderline behaviour disorder (Panagopoulos et al., 2013; Stoltenberg, Lehmann, Christ, Hersrud, & Davies, 2011; Vaags et al., 2012). The function of this gene has not been investigated in relation to tendon pathogenesis.

Overall, the analysis of the potential role of genes with top signal SNPs in tendon pathology identified three genes involved in the same pathway – the Wnt signalling

pathway. This pathway activates multiple transcriptional programs, which are essential in stem cell activation, cell proliferation and differentiation. The Wnt pathway has been investigated in association with various tissues, hemopoietic system and cancer (Clevers, 2006). The Wnt pathway also targets MMPs, which are involved in the homeostasis of the extracellular matrix in the tendon. Although genetic variations of MMPs have been investigated previously, regulatory mechanisms of their expression derived through the Wnt pathway have not been studied in relation to tendon pathology. Based on the findings of this study, future research of the Wnt pathway in the tendon may assist to a better understanding of the tendon pathogenesis.

4.4.2 Replication analysis of GWAS results

Complex traits, such as overuse injuries may have a very complicated interrelationship of genetic and other extrinsic factors. Therefore, replication of results is of high importance when such complex traits are investigated. At the time of publication, there has been only one previous GWAS of Achilles tendon injuries published, and this study failed to replicate results reported by candidate gene studies, which used a candidate gene approach (Kim et al., 2017). In the current study, we also attempted replication of findings from previous studies, which utilised a candidate gene approach and a cross-sectional replication of GWAS results between Kim et al. publicly available data. The results of the current GWAS differed from any previous findings of the candidate gene approach studies and the attempted replication of GWAS results from Kim et al., did not support their findings. Additionally, the most significant SNPs identified in this study were not found to be significant in the Kim et al data. However, the characteristics of populations and the cause of Achilles tendon injuries studied by Kim et al. and our research group differed. Our study was focused on Achilles tendon injuries that resulted from physical activity overload, specifically running. Kim et al. indicated that cases of Achilles tendon injuries in their study were of unknown nature and may not have been associated with physical activity. This was acknowledged as a limitation of their study and could explain discrepancies in the findings with our results and candidate gene approach studies, which predominantly used physically active people as cases. Putative replication of the rs12722 and rs1110495 SNPs was observed in our analyses, however, these results should be interpreted with caution

due to the imputed nature of the replicated data and low statistical significance. Further confirmation by direct genotyping is required to verify these findings.

This issue of failed replication of the results demonstrates the importance of consistency in the studied trait definition and characteristics, together with other critical requirements for successful GWAS. A large sample size is an essential requirement of studies seeking genetic associations with complex traits such as injuries. Additionally, the opportunity to perform a meta-analysis and expand the sample size may increase the chances of stronger associations.

4.4.3 Limitations of the study

A sample size with sufficient statistical power is critical to the success of GWAS to detect genetic associations for human complex diseases. The minimum required sample size for a case-control study with 80 power depends on multiple parameters: disease prevalence, genotype RR, MAF, LD and type of inheritance (Pfeiffer & Gail, 2003; Scherag, M ller, Dempfle, Hebebrand, & Sch fer, 2003). These factors are difficult to control and could be modulated by the employment of the sufficient sample size. The sample size calculation helps to avoid either larger sample size and therefore, additional costs, or smaller sample size and therefore, increased risk of false-negative rates (Hong & Park, 2012).

In the initial planning stages of this study, we aimed to recruit a minimum of 800 participants with each type of injury (cases) and 800 or more uninjured controls. These numbers were initially based on the popularity of this form of physical activity (Section 1.2) and the previously reported relatively high prevalence (10) of both injury types (Section 1.4.2 and Section 1.5.1). Using these sample sizes our initial power calculations (80 power; prevalence 10 ; $p<5x10^{-8}$) showed that common SNPs with a MAF 0.3 and an RR 1.5 would be within the significance threshold (Section 1.8.3). For comparison, the MAF of previously investigated polymorphisms varied between 0.05 and 0.5 (Table 4.6). However, in order to analyse SNPs with a lower MAF, for example, a MAF 0.05 at the same RR 1.5, the sample size of the study should be significantly increased and comprise approximately 5,000 participants in each group. Similarly, if RR 1.3 then the sample size must exceed 20,000 samples in each group to allow us to analyse SNPs with MAF<0.5. On the other hand, if RR 2

and a sample size is 1,000 participants in each group, then SNPs with MAF between 0.05 and 0.5 could be included in the analyses.

As previously discussed in Section 2.4.1, a 25-month extensive recruitment effort resulted in the engagement of 4,720 unique responses from recreational runners. However, only 35 (n 1,651) met the predetermined selection criteria (age, ethnicity, weekly running distance, and medical and health criteria) for inclusion in the genetic component of the study. As previously reported in Section 4.1.9, 70 (n 1,165) of eligible runners agreed to participate and provided a saliva sample. Quality control of the data eliminated 66 samples, with a total of 1,099 samples (94) included in the final analyses.

Despite our extensive efforts, sample numbers were considerably lower than the initial recruitment targets; 21.4 (171/800) with a prior AT injury and 95.9 (767/800) uninjured controls. Repeated power calculations of our final cases and control (170:770) recruitment showed that 80 power could only be reached for polymorphisms with average MAF 0.3 and a genetic RR of 2 and more (Figure 4.3). This high RR is very uncommon for polymorphisms associated with complex traits, such as overuse injuries. Genetic variants associated with multifactorial diseases such as type I and type II diabetes, cardiovascular disease, autoimmune, neuropsychiatric conditions and cancer have been successfully identified and replicated in diverse populations (Manolio, Brooks, & Collins, 2008). However, the majority of the identified polymorphisms had a modest effect with a risk (OR) varying between 1.2 and 1.5, and rarely exceeding a value of 2 (Manolio et al., 2008). This range of the OR is noticeably lower when compared, for example, to the risk of 3.2 identified in the association between apolipoprotein E ε4 allele and Alzheimer's disease (Rubinsztein & Easton, 1999). Importantly, utilisation of a GWAS and scanning of hundreds of thousands of SNPs also allowed identifying key pathways, which were not implicated previously in those multifactorial diseases. Therefore, mapping out genetic variants with relatively low risk and identification of their roles in the pathophysiology of multifactorial diseases still may contribute to the understanding of the diseases and search for preventive measurements and new therapeutic approaches.

Overall, the main limitation of this study, a small sample size, may be resolved by the employment of additional recruitment strategies and collaborations involving data

pooling and meta-analyses. Although over 4,700 recreational runners provided their detailed data for the epidemiological analyses, 65 of them were not eligible for the genetic study due to further selection criteria: in particular age limit, minimal weekly running distance, ethnicity, chronic conditions, which affect the musculoskeletal system, and smoking status. The upper age limit of 50 years and minimal weekly running distance were crucial factors for the study as the purpose of the project was to investigate genetic variants associated with overuse injuries acquired through regular physical activity. Therefore, these two factors excluded potential confounding factors associated with ageing and physical inactivity, predominantly chronic conditions, which may lead to similar diagnoses, but of different origin. Additionally, recreational runners were asked about injuries which occurred only in the past two years, but not lifetime injury history. This decision was made in order to avoid an increased risk of inaccuracies in the retrospective self-reported injury data (Gabbe et al., 2003).

The final GWAS sample size included approximately 170 cases of each type of injury and 767 uninjured controls, whereas original power calculations recommended the recruitment of at least five times more cases. Therefore, using the same selection criteria and recruitment strategies that did not target runners with the injuries of interest, 25,000 completed survey responses would be required to deliver the required number of cases of each injury. Based on the Australian population data and the popularity of running among Australians, there are approximately 900,000 recreational runners between 18 and 50 years of age (Australian Bureau of Statistics, 2015, 2016b). If the average prevalence of bone stress injuries and Achilles tendon injuries were 10 each, then potentially approximately 90,000 of Australian recreational runners experienced one of these injuries. If we were to recruit 800 cases of each type of injury, this would require only 1 of the injured runners in Australia. Although the ethnic background was one of the eligibility criteria, it was not a particularly restrictive factor. In our study, 90 , of recreational runners who completed the survey were of Caucasian European or Mediterranean background (Manzanero et al., 2018). Moreover, current methods of bioinformatics allow researchers to control for ethnicity as a potential confounding factor. Therefore, the ethnic background could be excluded from the eligibility factors in the future. These crude calculations demonstrate that there are potentially the numbers of study participants available in Australia, therefore,

it is feasible that the employed recruitment strategies were not ideal for recruitment of that large sample size. Of course, it is important to take into consideration usually predetermined time limitations and financial expenditures, which may limit recruitment efficiency.

An approach, which could increase the sample size, could be a collaboration with health practitioners who would encourage eligible runners to participate in the study. This could improve the quality of the collected injury data as participants injuries would be diagnosed by a professional and therefore decrease proportions of self-diagnoses. In addition, potential collaboration with medical professionals could give access to databases with lifetime diagnoses of research interest, hence the criterion of the injury report time limit of two years could be excluded. However, this approach would require expansion of communications and promotion of the project in the professional medical community, significant relationship development and a willing group of medical professionals with the time and capacity to be involved.

A critical assessment of the sample size in multiple GWAS demonstrated that the optimal size is crucial for the achievement of the statistical power of 80 . In situations where the number of cases is limited, then 80 power can still be achieved by increasing the relative proportion of case-controls from 1:1 to 1:4 case: controls. Thus, in the study, a proportion of 600:2400 would allow us to investigate SNPs with MAF 0.3 and RR 1.5. Interestingly, 1:4 was the final proportion of cases and controls in this study (170:770), which supports a good chance of recruitment in the proposed proportion in the future.

The final approach to substantially increasing a sample size is a collaboration with other research groups that collected similar phenotypic and genetic data. Certainly, collaboration and data pooling increase the chance of producing high-quality research based on a large sample size. Development of online data-sharing platforms, biobanks and meta-analyses makes collaboration more achievable and increases chances to identify genetic variants at the required significance level. Currently, the methodology of meta-analyses allows researchers to merge data. However, prior to data sharing and pooling some important factors should be taken into consideration: heterogeneity of data, differences in the injury definitions and diagnoses, selection criteria and

utilisation of different genotyping platforms and imputation software (Evangelou & Ioannidis, 2013).

Collection of a sample size with sufficient statistical power is a common challenge for GWAS. It is crucial to calculate the correct sample size during the study design and consider all contributing factors such as disease prevalence, relative risk, and allele frequencies. Access to large patient databases and collaboration are the key solutions in the recruitment of a large sample size.

5. Chapter Five – Genome-wide association study of genetic variants associated with bone stress injuries

Addendum

Contributions to Chapter 5:

Mariia Kozlovskaia:

- Sample collection and registration
- DNA extraction
- GWAS data analysis
- Imputed data analysis
- Author of the chapter

Staff members of the Translational Research Institute:

- GWAS analysis

Paul Leo:

- Data imputation
- Assistance with GWAS data analysis

Kevin Ashton:

- RNAseq data analysis
- Assistance with GWAS data interpretation
- Editing of the chapter

Rebecca Grealy:

- Sample registration
- DNA extraction

Nicole Vlahovich:

- Editing the chapter

5.1 Introduction

Bone stress injuries develop as a result of the repetitive loading of the bone when intensive weight-bearing activity is involved, particularly marching, jogging and running (Mattila et al., 2007). Bone stress injuries usually begin as stress reactions, which may develop into stress fractures and then complete bone fractures (Warden et al., 2014). Typically, bone stress injuries associated with loading of lower limbs develop in the tibial shaft and metatarsal bones, whereas upper leg and knee are less common locations of these injuries (Mattila et al., 2007). According to a systematic review, one of the most prevalent running-related injuries was a specific bone stress injury - medial tibial stress syndrome (MTSS) (Dias Lopes et al., 2012). Risk factors which contribute to the development of bone stress injuries may be divided into two groups: factors modifying the load applied to a bone (anatomic and biomechanical characteristics, training habits), and factors affecting bone density and its ability to resist the load (sex, age, hormonal status, chronic conditions, diet, genetic factors) (Warden et al., 2014).

Several factors affecting the load applied to a bone were associated with an increased risk of bone stress injuries. Abnormal static alignment and leg length discrepancy may affect movement patterns during a running session and contribute to the development of bone stress injuries (Bennell et al., 1999; Warden et al., 2014). The rapid increase in training load may lead to accumulation of microdamage in the bone and development of the injury, especially if the base level of physical conditioning was low (V lim ki et al., 2005; Warden et al., 2014). In contrast, a gradual increase in the load and training regimen with rest periods allow the bone to adapt to the load and decrease the risk of bone stress injuries (Bennell et al., 1999). A higher incidence of bone stress injuries in females was reported by several studies (Mattila et al., 2007; Newman et al., 2013; Protzman & Griffis, 1977) and also observed in the project's recruited cohort (Chapter 3). Some of the factors that affect bone density are associated with sex. These factors, associated with female sex, included low bone mineral density (BMD), menstrual disturbances (hormonal status disruptions) and nutritional issues, leading to a negative energy balance (Korsten-Reck, 2011). Then a new term 'Relative Energy Deficiency in Sport' was introduced, and all these factors became applicable to both sexes (Mattila et al., 2007; Mountjoy et al., 2015). Increased BMI and a higher proportion of body fat were associated with bone stress injuries (Beck et al., 2015; Newman et al., 2013). Additionally, during childhood and teenage, the bone may be

more vulnerable to injuries due to its immature condition and development processes during those periods. Bone density also decreases with age, which may lead to a higher risk of bone stress injuries later in life. Therefore, age may be an important risk factor, especially when hormonal status and training loads are considered as co-factors (Bennell et al., 1999).

Two studies of bone stress injuries in monozygotic twins suggested genetic input to the development of bone stress injuries (Singer et al., 1990; muda et al., 2011). In addition, another study on recurrence of these injuries in athletes also contributed to the investigation of genetic variants, which may contribute to the development of bone stress injuries (Bennell et al., 1996b). It was proposed that genetic variants associated with bone stress injuries would most likely be those that have previously been associated with BMD, bone formation and chronic disorders affecting bone strength, such as osteoporosis. Calcium and vitamin D are essential for bone mineralisation, and supplementation of calcium and vitamin D is an efficient means of bone stress injury prevention (Lappe et al., 2008). Using a candidate gene approach several polymorphisms in the vitamin D receptor (*VDR*) gene were found to be associated with changes in bone mineral density (Ferrari et al., 1998; Kehoe & Montgomery, 2006). Additionally, polymorphisms in the *VDR* gene and a gene encoding bone morphogenetic protein 2 (*BMP2*) were associated with osteoporosis (Raisz, 2005; Styrkarsdottir et al., 2003). Also, polymorphisms in the *VDR* gene and a calcitonin receptor gene (*CALCR*) were significantly associated with stress fractures in military personnel (Yanovich et al., 2012). The Wnt pathway has previously been shown to play an essential role in bone formation, with disruptions in this pathway linked to osteoporosis (Piters et al., 2008). Hence, investigation of polymorphisms in genes encoding proteins involved in the Wnt pathway, and proteins antagonising this pathway, was undertaken using a candidate gene approach. This study showed a significant association between polymorphisms in the LDL receptor-related protein 5 (*LRP5*) and sclerostin (*SOST*) and changes in bone mineral density (Piters et al., 2008). The RANK/RANKL/OPG pathway plays an important role in bone remodelling and adaptation. Genes encoding proteins in this pathway were examined, and polymorphisms in these genes were shown to be significantly associated with stress fractures in elite athletes (Varley et al., 2015). The same cohort of elite athletes was used to investigate polymorphisms in the gene encoding P2X7 receptor, which is

involved in osteoblast and osteoclast response to stress. This study identified two significant polymorphisms associated with stress fracture prevalence in elite athletes (Varley et al., 2016).

Overall, the knowledge of genetic predisposition to bone stress injuries is limited to just a few studies and requires deeper research involving the recruitment of specific cases with bone stress injuries and employment of less biased molecular methods such as GWAS (as discussed in Chapter 4). Whilst at the time of publication, there are no GWAS focusing on bone stress injuries, several related studies have attempted to localise genetic markers associated with BMD and bone fractures. One large study included three different populations from Iceland, Denmark and Australia and identified four genomic regions associated with BMD: *ZBTB40–WNT4, RANKL*, osteoprotegerin (*OPG*) and estrogen receptor 1 (*ESR1*) (Styrkarsdottir et al., 2008). A large meta-analysis of 17 GWAS investigating genetic markers in association with BMD and identified 32 novel genomic loci associated with BMD, involving multiple Wnt pathway factors (e.g. *SOST, LRP4, WNT4, DKK1*), a pathway, which is involved in mesenchymal cell differentiation (e.g., *RUNX2, SOX4, SOX9*), the RANK/RANKL/OPG pathway, and a pathway, which is responsible for foetal development of the skeleton (e.g. *SPP1, ME2C, RUNX2, SOX6*) (Estrada et al., 2012). Additionally, several genetic variants were identified as important factors for factures associated with BMD. This meta-analysis outlined a highly polygenic nature of genetic variation of BMD and bone fractures. Therefore, this part of the project aimed to identify novel polymorphisms associated with bone stress injuries in recreational runners.

5.2 Methods

5.2.1 Participants

Selection criteria for eligible participants, as previously described in Section 4.2.1.

5.2.2 Sample collection

Sample collection, as previously described in Section 4.2.2.

5.2.3 DNA Extraction

DNA extraction protocol, as previously described in Section 4.2.3.

5.2.4 Genotyping and Quality Control

Quality control of the genomic data was performed on all collected samples prior to their assignment to cases and controls, as previously described in Section 4.2.4.

5.2.5 Statistical Analyses

Statistical analyses of cases with reported bone stress injuries and uninjured controls, as previously described in Section 4.2.5.

5.2.6 Imputation

Imputation of additional genotypes as previously described in Section 4.2.6.

5.2.7 Replication Analysis

Previously identified SNPs described in Table 1.2 were further investigated to attempt replication of those results in the current study, and seeking identified significance levels for previously described SNPs in the imputed data.

5.3 Results

5.3.1 Phenotypic characteristics of the participants

5.3.1.1 Statistical comparison of BS genomic cohort and BS injury cohort

Recruitment resulted in 4,720 unique responses from recreational runners. After the application of the predetermined filters, a bone stress (BS) injury cohort was formed and comprised 475 bone stress injury cases and 1969 uninjured controls, resulting in a total of 2,444 recreational runners. Out of 1,165 participants recruited for the genetic arm of the study, 1,099 passed all required quality control steps and comprised 767 uninjured runners and 174 runners with bone stress injuries, resulting in a total of 941 runners, which were included in the BS genomic cohort. The remainder of the genotyped runners reported Achilles tendon injuries and were analysed in Section 4.3.

The physical characteristics and reported ethic background of the BS genomic cohort are displayed in Table 5.1, and the physical characteristics of the BS injury cohort are displayed in Table 3.9.

Injury cases comprised 18.5 of BS genomic cohort, and this rate was similar to the rate of cases in the BS injury cohort of 19.4% (χ^2_1 0.39, p 0.52). The BS genomic cohort had a significantly higher rate of females than BS injury cohort (59.5 versus 56.5%, χ^2_1 3.35, p 0.04). However, the compared cohorts did not differ by the rates of injured males and females (χ^2_1 1.27, p=0.18; χ^2_1 2.26, p 0.08, respectively). Moreover, the rates of injured males and females were similar in both BS genomic and BS injury cohorts (17 and 19.5 in males, 18.6 and 20.1 in females). The BS genomic cohort also contained a significantly higher rate of runners with normal BMI than the BS injury cohort (77.7 versus 74.4 , χ^2_1 3.87, p 0.03).

Participant age (between 18 and 50 years) was a specific selection criterion for the genomic component of the study. The BS injury cohort comprised 17.7 of runners aged over 50 years old. These older runners were excluded from the remaining analysis, and four age groups were formed: 18-24, 25-34, 35-44 and 45-50. The statistical comparison of BS injury cohort and BS genomic cohort by the age group distribution showed a significant difference between the two cohorts (χ^2_3 13.58, p 0.004). The uninjured groups of BS injury cohort and BS genomic cohort also were statistically different by age groups (χ^2_3 12.23, p 0.006). However, age group distributions were not statistically different between injured groups of BS injury cohort

and BS genomic cohort (χ^2_3 1.59, p 0.65). The lower rates of younger runners in the BS genomic cohort than in the BS injury cohort could be explained by the unwillingness of younger runners to provide a sample for the genetic study.

The reported ethnic background of four grandparents was a key criterion for the genetic arm of the study. A runner who reported at least 75 of Caucasian European or Mediterranean background were eligible for genetic analysis. Thus, 87.6 of BS injury cohort reported 100 or a combination of these two backgrounds. Among the runners who were included in the genetic analysis, 94.3 were of Caucasian European background, and the remaining 5.7 had a combination of backgrounds with a prevalence of Caucasian European and Mediterranean backgrounds.

When BS genomic cohort was compared to BS injury cohort by training characteristics of recreational runners, there was no significant difference by weekly running distance (χ^2_1 1.98, p 0.37), number of running sessions per week (χ^2_3 0.79, p 0.85), running terrains (χ^2_4 1.94, p=0.74), participation in other sports besides running (χ^2_1 1.91, p=0.16), stretching in relation to running sessions (χ^2_1 0.14, p 0.7), and wearing orthotics (χ^2_1 0.1, p 0.75). The two compared cohorts were significantly different by years of running experience (χ^2_3 177.46, p<0.001). However, when injured and uninjured groups of BS injury and BS genomic cohorts were compared independently, injured cohorts were not significantly different by years of running experience (χ^2_3 3.86, p=0.28), whereas uninjured groups differed significantly (χ^2_3 15.72, p 0.001). The BS genomic cohort also significantly differed from BS injury cohort by the reported race pace (χ^2_4 14.48, p 0.005). However, injured groups of these two cohorts did not differ significantly by the race pace groups (χ^2_4 2.04, p 0.72), whereas uninjured groups significantly differed by this variable (χ^2_4 13.3, p 0.009). In summary, BS genomic cohort differed from BS injury cohort by proportions of females, BMI group distribution, frequencies of age groups, reported race pace and years of running experience.

5.3.1.2 Statistical comparison of injured and uninjured runners within the BS genomic cohort

The BS genomic cohort consisted of 564 females (59.9) and 377 males (40.1). Among these 941 runners, the proportions of injured females (n 110, 19.5) and males (n=64, 17%) were similar (χ^2_1 0.96, p 0.33). The frequencies of age groups were significantly different between injured and uninjured groups (χ^2_3 11.1, p 0.01), with higher rates of runners aged 18-34 in the injured group than in the uninjured group (42.5 versus 31.6). However, injured and uninjured groups did not differ by frequencies of BMI categories (χ^2_3 0.43, p 0.93).

Table 5.1 Physical characteristics of runners with bone stress injuries and uninjured runners in the BS genomic cohort.

Physical characteristics		Runners with BS injuries (*N*=174)		Uninjured Runners (*N*=767)	
		n	***%***	***n***	***%***
Sex	Male	64	36.8	313	40.8
	Female	110	63.2	454	59.2
Age group*	18-24 years	18	10.3	38	5
	25-34 years	56	32.2	204	26.6
	35-44 years	71	40.8	371	48.4
	45-50 years	29	16.7	154	20.1
Body mass index	Underweight (<18.5 kg/m^2)	5	2.9	16	2.1
	Normal (18.5 to <25 kg/m^2)	135	77.6	596	77.7
	Overweight (25 to <30 kg/m^2)	31	17.8	141	18.4
	Obese (≥30 kg/m2)	3	1.7	14	1.8
Reported ethnic background	100 CE	159	91.4	728	94.9
	75 CE + 25 MT/Other	11	6.3	14	1.8
	50 CE + 50 MT/Other	4	2.3	19	2.5
	100 MT	0	0.0	5	0.7
	75 MT + 25 Other	0	0.0	1	0.1

BSI – bone stress injury, CE – Caucasian European, MT – Mediterranean, * - statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.05).

Training habits of injured and uninjured runners included in the BS genomic cohort are described in Table 5.2. There were significantly more runners who reported running over 40 km per week in the injured group than in the uninjured group (36.2 versus 26%, χ^2_2 8.02, p 0.02). The injured and uninjured groups also differed by frequencies of the reported years running experience (χ^2_3 12.26, p 0.007). The injured group had approximately half the proportion of inexperienced runners (≤2 years) than the uninjured group (11.5 versus 22.4). The compared groups did not differ by the frequencies of running sessions per week (χ^2_3 2.59, p=0.46), race pace (χ^2_4 3.56, p=0.46), preferred running terrain (χ^2_6 11.49, p 0.07), participation in other sports (χ^2_1 0.02, p=0.88), and stretching in relation to running sessions (χ^2_1 2.96, p 0.09). However, injured runners were more likely to report wearing orthotics (33.3 versus 14.1%, χ^2_1 36.6, p<0.001).

Overall, injured and uninjured groups differed significantly by the frequencies of age groups, reported weekly running distance, years of running experience and wearing orthotics.

Table 5.2 Training characteristics of runners with bone stress injuries and uninjured runners in the BS genomic cohort.

Training characteristics		Runners with BS injuries (*N*=174)		Uninjured Runners (*N*=767)	
		n	***%***	***n***	***%***
Weekly running distance*	15-20	35	20.1	205	26.7
	20-40	76	43.7	362	47.2
	40+	63	36.2	200	26
Running experience*	≤2 years	20	11.5	172	22.4
	3-5 years	59	33.9	196	25.6
	6-9 years	32	18.3	127	16.6
	10+ years	63	36.2	271	35.3
Running sessions per week	1 session	0	0	2	0.3
	2 or 3 sessions	66	37.9	327	42.6
	4 or 5 sessions	85	48.9	358	46.7
	6+ sessions	22	12.6	74	9.6
	N/A	0	0	6	0.8
Race pace	<4 min/km	25	14.4	82	10.7
	4-5 min/km	66	37.9	278	36.3
	5-6 min/km	64	36.8	291	37.9
	6-7 min/km	15	8.6	95	12.4
	7 min/km	4	2.3	18	2.3
	N/A	0	0	3	0.4
Running terrain	Bitumen	78	44.8	322	42
	Cement	48	27.6	256	33.4
	Grass	8	4.6	16	2.1
	Hard dirt/gravel	35	20.1	146	19
	Treadmill	2	1.1	20	2.6
	Sand	0	0	4	0.5
	Synthetic	3	1.7	3	0.4
Participation in other sports	Yes	132	75.9	583	76
	No	40	23	182	23.7
	N/A	2	1.1	2	0.3
Stretching	Yes	114	65.5	447	58.3
	No	60	34.5	318	41.5
	N/A	0	0	2	0.3
Orthotics**	Yes	58	33.3	108	14.1
	No	115	66.1	658	85.8
	N/A	1	0.6	1	0.1

BS – bone stress, CE – Caucasian European, MT – Mediterranean, * - statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.05), **- statistically significant difference between injured and uninjured groups within the AT genomic cohort (p<0.001).

5.3.1.3 Discussion of phenotypic characteristics of investigated cohorts

Bone stress injuries are one of the most common running-related injuries, with annual incidence rates varying between 4-5 among adolescent runners and over 20 in elite collegiate runners (Tenforde, Kraus, & Fredericson, 2016). Previous studies of bone stress injuries in runners identified a higher incidence rate in female runners and female military conscripts than in male counterparts (Mattila et al., 2007; Newman et al., 2013). In this study, the two-year incidence of bone stress injuries in the large injury cohort, described in Chapter 3 (N 4,284) was 11.1 and was significantly higher in female runners than in male runners (12% versus 10%, χ^2_1 4.19, p 0.04). However, after the exclusion of runners reported other types of injuries and formation of BS injury and BS genomic cohorts, bone stress injury rates were similar for males and females in both analysed cohorts. This may be an important observation because several risk factors of bone stress injuries were originally linked to female sex and a systematic review of MTSS showed an association between female sex and increased risk of an injury (Newman et al., 2013). However, it is important to acknowledge that male runners may be as prone to bone stress injuries as female runners and may be affected by the same risk factors, such as changes in hormonal status, dietary disorders and low bone mineral density (Mountjoy et al., 2015).

Since age was an ambiguous factor and both adolescents and elderly people are prone to develop bone stress injuries, these age groups were excluded from the BSI genomic cohort. Observed statistical difference between the described cohorts by the age group distribution, particularly between uninjured subgroups, could be explained by the lower interest of younger people to participate in the genetic arm of the study (Manzanero et al., 2018).

5.3.2 Genetic polymorphisms associated with bone stress injuries

Following QC and filtering of the SNP arrays, logistic regression analysis was used to calculate p-values across 281,168 genetic polymorphisms. Firstly, observed and expected p-values were compared, the QQ plot showed that these values were similar, and none of the genotyped polymorphisms reached a statistical significance of $p<5x10^{-8}$ (Figure 5.1). Obtained p-values were then visualised on a Manhattan plot across the 22 autosomal chromosomes (Figure 5.2). None of the genotyped SNPs

surpassed either the genome-wide (p<5x10^{-8}) or suggestive (p<10^{-5}) significance thresholds. We, therefore, decided to identify the top 20 most significant SNPs, which required the significance threshold to be reduced to p<9.4x10^{-5}. As shown in Table 5.3, using this lower threshold, the 20 most significant SNPs were predominantly located on chromosomes 5, 6 and 9.

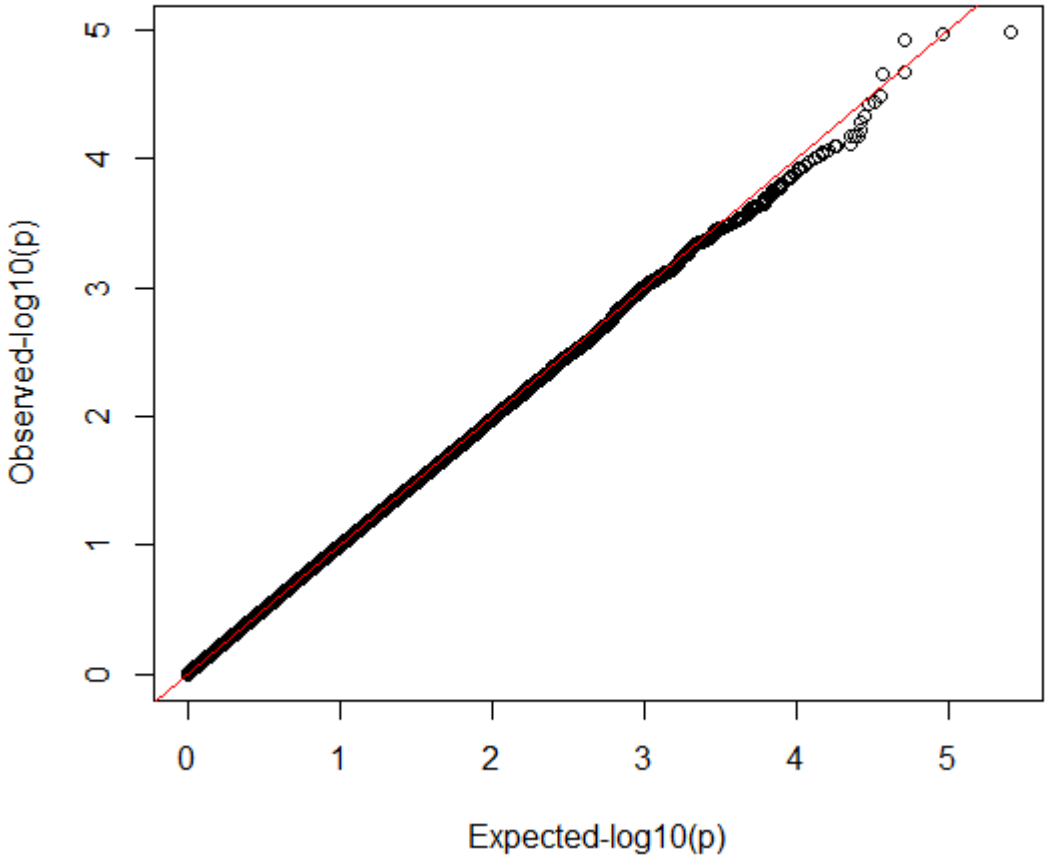


Figure 5.1 QQ plot for the bone stress injury of observed and expected *p*-values, -log10 transformed.
The plot shows observed *p*-values plotted on the y-axis (black dots) and theoretical *p*-values plotted on the x-axis (red line). The black dots align closely with the red line indicating that there is little to no statistical difference between the BS case and control groups.

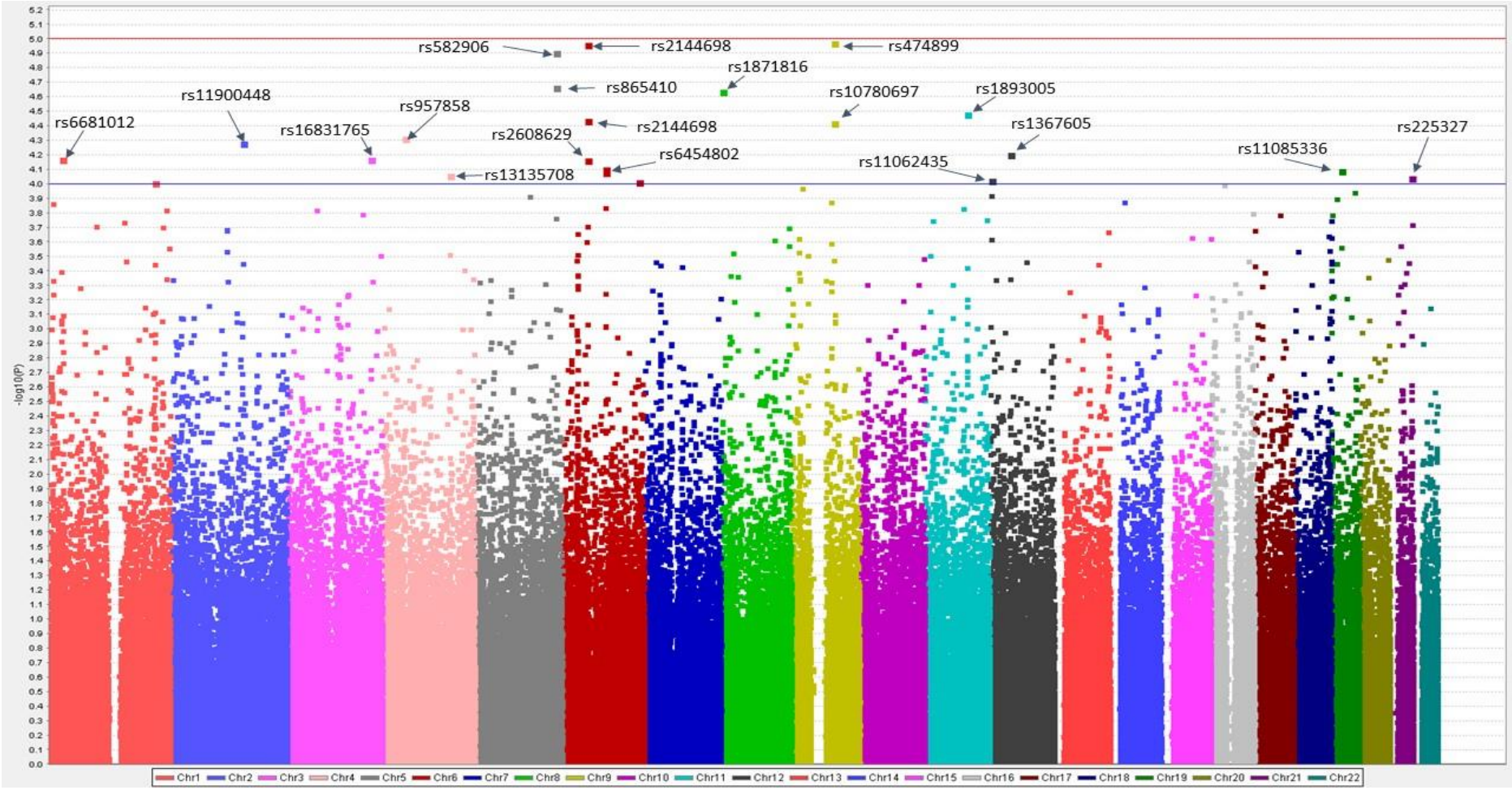


Figure 5.2 Manhattan plot of *p*-values calculated for the bone stress injury case-control analysis, -log-transformed.
Y-axis shows $-\log_{10}$ *p*-value for association with a bone stress injury. The red line indicates the suggestive significance threshold of $p<1x10^{-5}$, the blue line indicates the threshold of $p<9.4x10^{-5}$, which identifies the top 20 most significant SNPs. Chromosome colour-coded legend is located along the x-axis. The rs numbers for each of the top 20 SNPs are indicated on the plot.

Table 5.3 Top-20 most significant genotyped SNPs, ordered by chromosomes and base-pair location.

SNP	Chr	BP	MAF	Gene	A1	OR	SE	*p*-value
rs6681012	1	25188064	0.29	*LOC105376876*	A	1.64	0.12	6.62E-05
rs11900448	2	149942519	0.45	*LYPD6B*	A	1.64	0.12	5.16E-05
rs16831765	3	173620010	0.19	*NLGN1*	G	0.46	0. 2	6.63E-05
rs957858	4	45737256	0.47	*Intragenic*	C	1.7	0.13	4.75E-05
rs13135708	4	139464920	0.3	*Intragenic*	A	1.75	0.14	8.55E-05
rs865410	5	167905111	0.24	*Intragenic*	C	1.77	0.13	2.12E-05
rs582906	5	167916628	0.48	*RARS*	A	1.71	0.12	1.23E-05
rs2180314	6	52617731	0.46	*GSTA2*	G	1.67	0.13	3.61E-05
rs2144698	6	52623807	0.13	*GSTA2*	A	1.75	0.13	1.08E-05
rs2608629	6	52625794	0.22	*GSTA2*	G	1.65	0.13	6.78E-05
rs6454802	6	90814199	0.23	*BACH2*	A	1.63	0.12	7.80E-05
rs11757155	6	90941240	0.18	*BACH2*	A	1.63	0.12	8.14E-05
rs1871816	8	3666456	0.13	*CSMD1*	A	4.03	0.33	2.25E-05
rs10780697	9	87654393	0.4	*LOC105376118*	G	1.7	0.13	3.73E-05
rs474899	9	87660797	0.32	*Intragenic*	A	1.76	0.13	1.05E-05
rs1893005	11	87006224	0.43	*TMEM135*	G	0.55	0.16	3.25E-05
rs11062435	12	3076665	0.39	*TEAD4*	G	1.79	0.15	9.35E-05
rs1367605	12	41684389	0.23	*PDZRN4*	G	1.71	0.13	6.14E-05
rs11085336	19	20478535	0.34	*LOC105372315*	A	1.67	0.13	8.01E-05
rs225327	21	43763206	0.36	*Intragenic*	A	1.67	0.13	8.99E-05

SNP – single nucleotide polymorphism, Chr – chromosome, BP – base pairs, MAF – minor allele frequency, A1 – effect allele, OR – odds ratio, SE – standard error, P – *p*-value.

Chromosome 6 contained the highest density of significant SNPs, with three SNPs located in the *GSTA2* gene (rs2180314, rs2144698, rs2608629) and two SNPs in the *BACH2* gene (rs6454802, rs11757155). The remaining 13 SNPs were distributed across 11 different chromosomes, with seven of them located in the protein-coding genetic regions. Two SNPs were located in uncharacterised loci (*LOC*), and four SNPs belonged to intragenic regions, with the most statistically significant SNP (rs474899) located in an intragenic region of chromosome 9. However, as previously discussed in Section 4.3.4 this data should be interpreted with extreme caution as the use of a reduced significance threshold ($p<9.4x10^{-5}$) greatly increases the probability of type I errors (statistically in the order of 28 false positives). In order to help control for this, the same approach from Chapter 4 was applied. Specifically, imputed data were utilised to support and identify those SNPs that were more likely to be associated with BSI and also to exclude unreliable signals.

5.3.3 Imputation of additional SNPs and visualisation of GWAS results

Imputation made available 23,699,623 additional genotypes to support the case-control analysis and visualisation of the array genotyped results. Case-control analysis for the imputed data was implemented under the same criteria as for the array genotyped data and resulted in calculated *p*-values for 7,438,753 genotypes. The imputed data was then visualised using Locus oom plots for the 20 previously described genotyped SNPs.

Five of the selected top 20 SNPs were located on chromosome 6, three of these mapped to the *GSTA2* gene (Figure 5.3A). Supporting imputed SNPs in *GSTA2* were distributed across surrounding genes encoding other members of glutathione S-transferase family: *GSTA7P*, *GSTA5* and *GSTA1*. However, only a few of imputed SNPs were in high LD (r^2 0.8). Two other SNPs on chromosome 6 mapped to the *BACH2* gene (Figure 5.3B). Multiple imputed SNPs with high LD were located around these two signals and spread along the *BACH2* gene. One of the genotyped SNPs (rs11757155) was surrounded by multiple supporting imputed SNPs in high LD. Therefore, due to the appearance of several signals in *GSTA2* and *BACH2*, both genes were considered for discussion regarding their potential function in the risk of bone stress injuries.

Six of the remaining 20 genotyped SNPs appeared in peak shaped LD patterns with other supporting imputed SNPs (Figures 5.4 and 5.5). However, not many genotyped SNPs had supporting imputed SNPs in high LD. Moreover, only three of these six SNPs were located within genes, specifically *LYPD6B* (rs11900448), *CSMD1* (rs1871816) and *TEAD4* (rs11062435). These three genes were included in the subsequent discussion of the potential link between their function and development of bone stress injuries. The other three SNPs were located in the uncharacterised or intragenic regions: rs6681012, rs11085336 and rs225327. Although, rs6681012 was located between two genes – *CLIC4* and *RUNX3*, none of the imputed SNPs belonged to either of these genes. Similarly, the closest to rs11085336 coding region was *MIR1270*, however, none of the supporting imputed SNPs were located in this gene. Finally, rs225327 was surrounded by highly conservative family of *TFF* (trefoil factor) genes, which are expressed predominantly in the gastrointestinal tissue. Hence, these three SNPs in the intragenic regions were excluded from the subsequent discussion.

One of two closely located signals on chromosome 5, rs582906 was located within the *RARS* gene (Figure 5.6C). The second genotyped SNPs (rs865410) was located in the intragenic region. Supporting multiple SNPs were widely distributed across neighbour genes, however, the majority of the imputed SNPs were in LD of r^2<0.6, and only several SNPs were clustered around these two genotyped SNPs. Similarly, widely scattered imputed SNPs were observed on the plots for rs16831765, located on chromosome 3 in the *NLGN1* gene, rs1367605 on chromosome 12 in the *PDZRN4* gene, and rs957858 on chromosome 4 in the intragenic region (Figure 5.6A, B and D). SNPs located in *NLGN1* and *PDZRN4* genes were supported by multiple imputed SNPs in high LD. Hence, these two genes and the *RARS* gene were included in the following discussion of their potential role in bone stress injury development.

Three of the remaining four signals rs13135708 (Figure 5.7A), rs474899 and rs10780697 (Figure 5.7B) were located in the uncharacterised and intragenic regions. Although the first SNP had multiple imputed surrounding SNPs in high LD, not of them were located in the surrounding genes. Yet the latter two SNPs were in proximity to the *NTRK2* gene, and some of the imputed SNPs were located in this gene. The last included in top 20 list signal rs1893005 belonged to the *TMEM135* gene on chromosome 11. This SNP was poorly supported by the imputed SNPs as the majority of them were in LD of r^2<0.6 (Figure 5.7C). These two genes – *NTRK2* and *TMEM135*

were included for discussion, *NTRK2* was in proximity of two identified signals and *TMEM135* has previously been investigated in relation to osteoporosis (Scheideler et al., 2008).

In summary, ten genes were included in the following exploratory investigation of their functions and potential involvement in the development of bone stress injuries: *LYPD6B*, *CSMD1, TEAD4, GSTA2, BACH2, RARS, NLGN1, TMEM135, PDZRN4* and *NTRK2*.

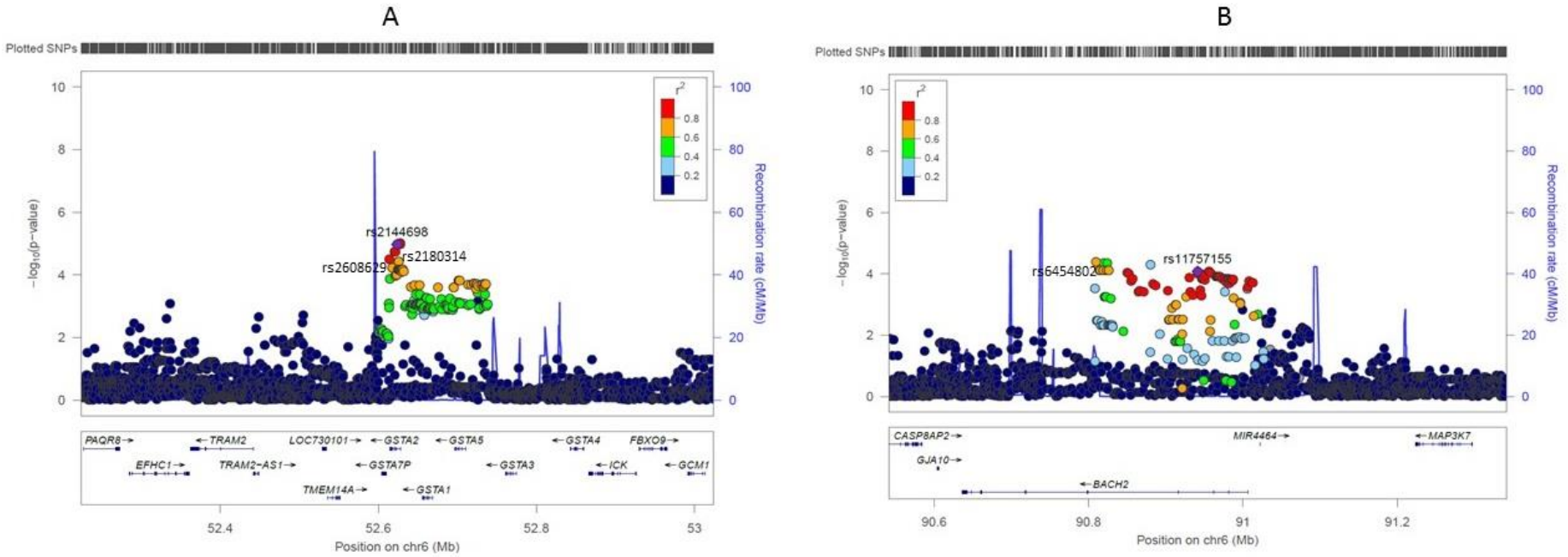


Figure 5.3 Locus Zoom plots for identified significant SNPs for bone stress injuries from chromosome 6.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals in *GSTA2* gene; B) signals in *BACH2* gene.

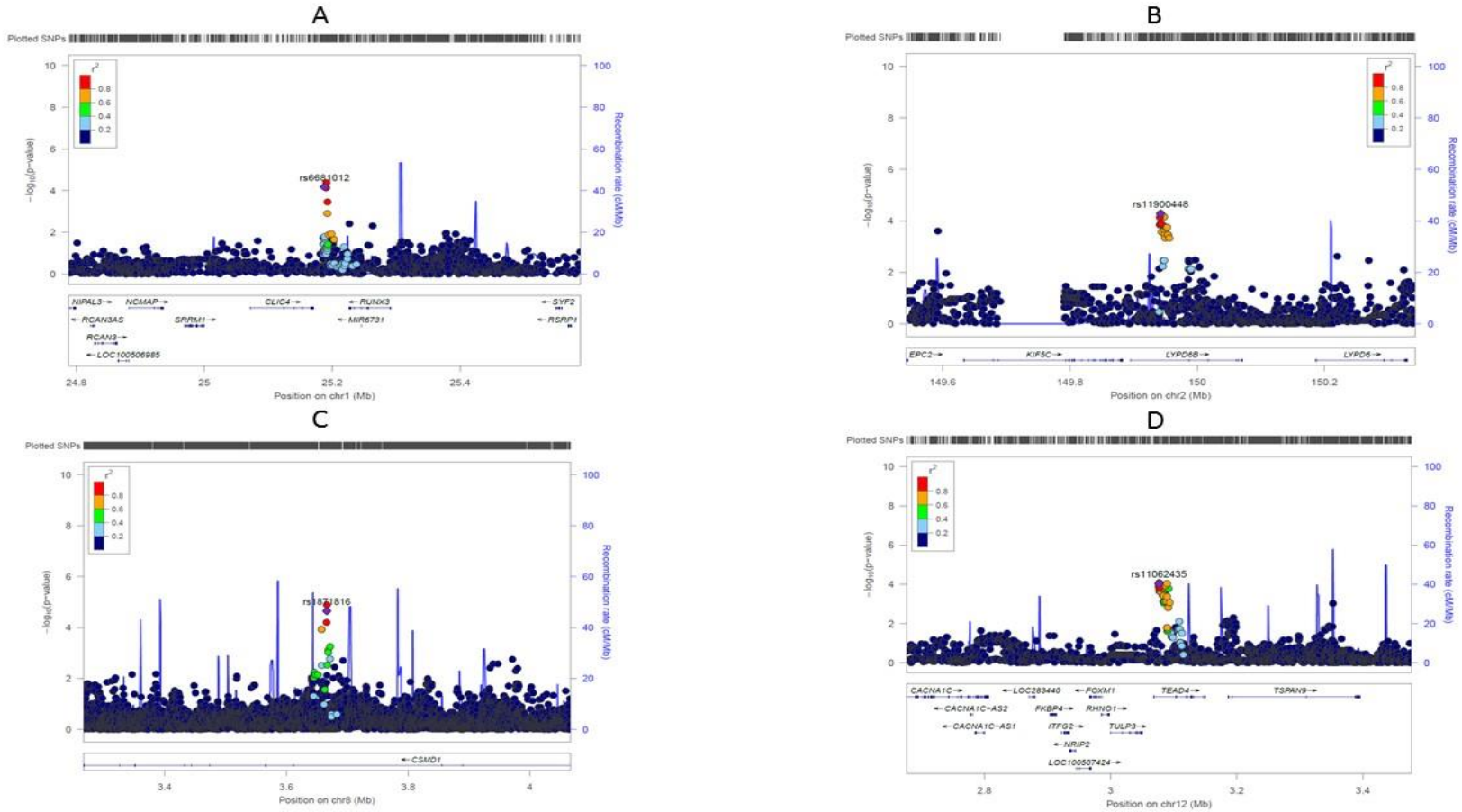


Figure 5.4 Locus Zoom plots for identified significant SNPs for bone stress injuries from chromosomes 1, 2, 8 and 12. Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 1 in the intragenic region; B) signals on chromosome 2 in *LYPD6B* gene; C) signals on chromosome 8 in *CSMD1* gene; D) signals on chromosome 12 in *TEAD4* gene.

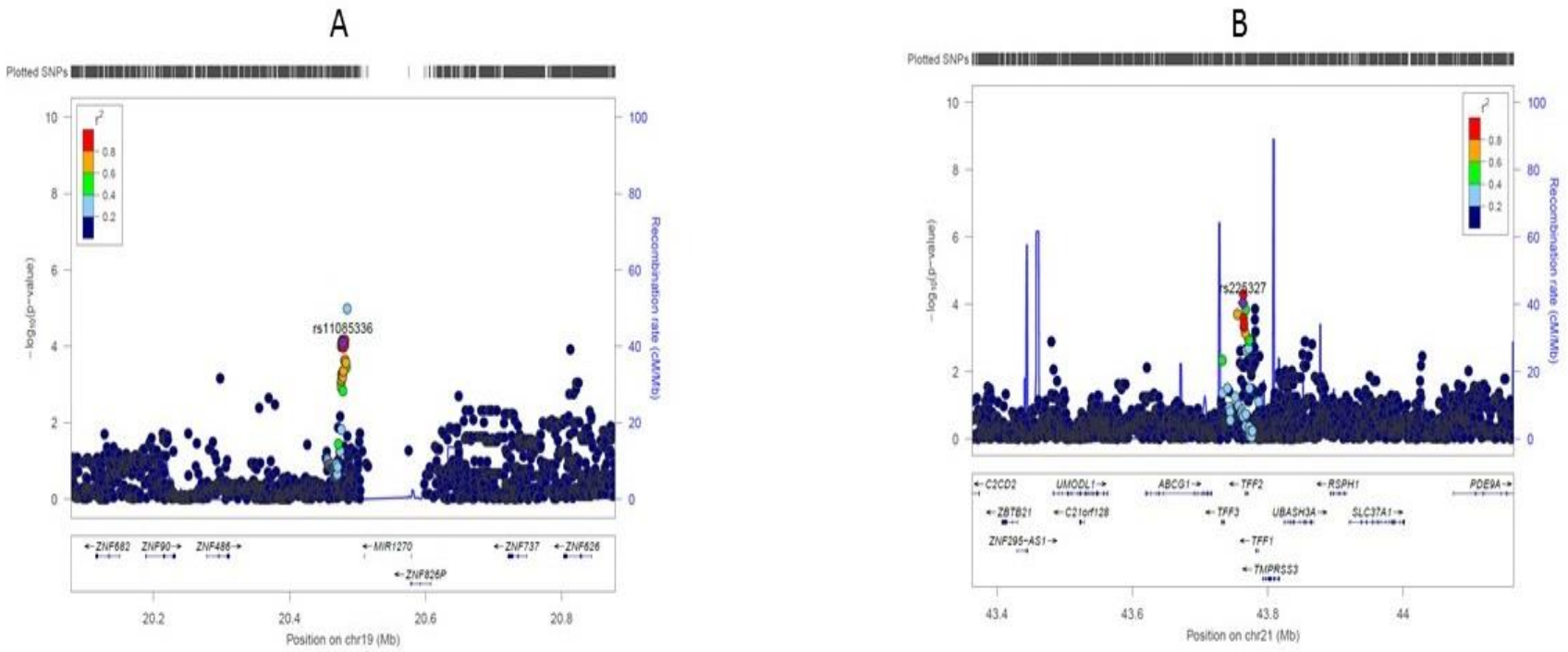


Figure 5.5 Locus Zoom plots for identified significant SNPs for bone stress injuries from chromosomes 19 and 21.
Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Regions where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 19 in the intragenic region; B) signals on chromosome 21 in the intragenic region.

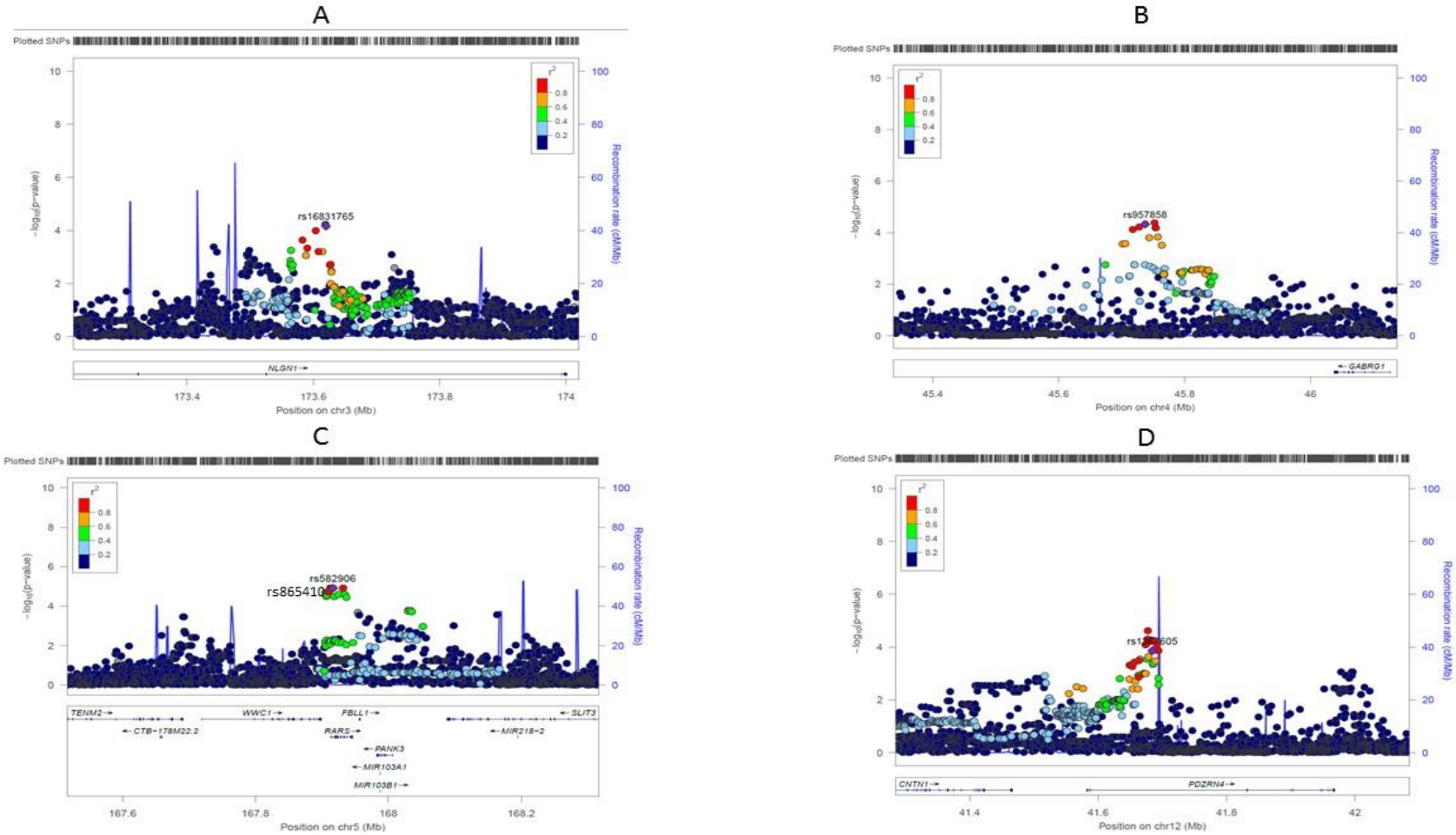


Figure 5.6 Locus Zoom plots for identified significant SNPs for bone stress injuries from chromosomes 3, 4, 5 and 12. Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 5 in the *RARS* gene; B) signals on chromosome 5 in the intragenic region; C, D) signals on chromosome 6 in *BACH2* gene.

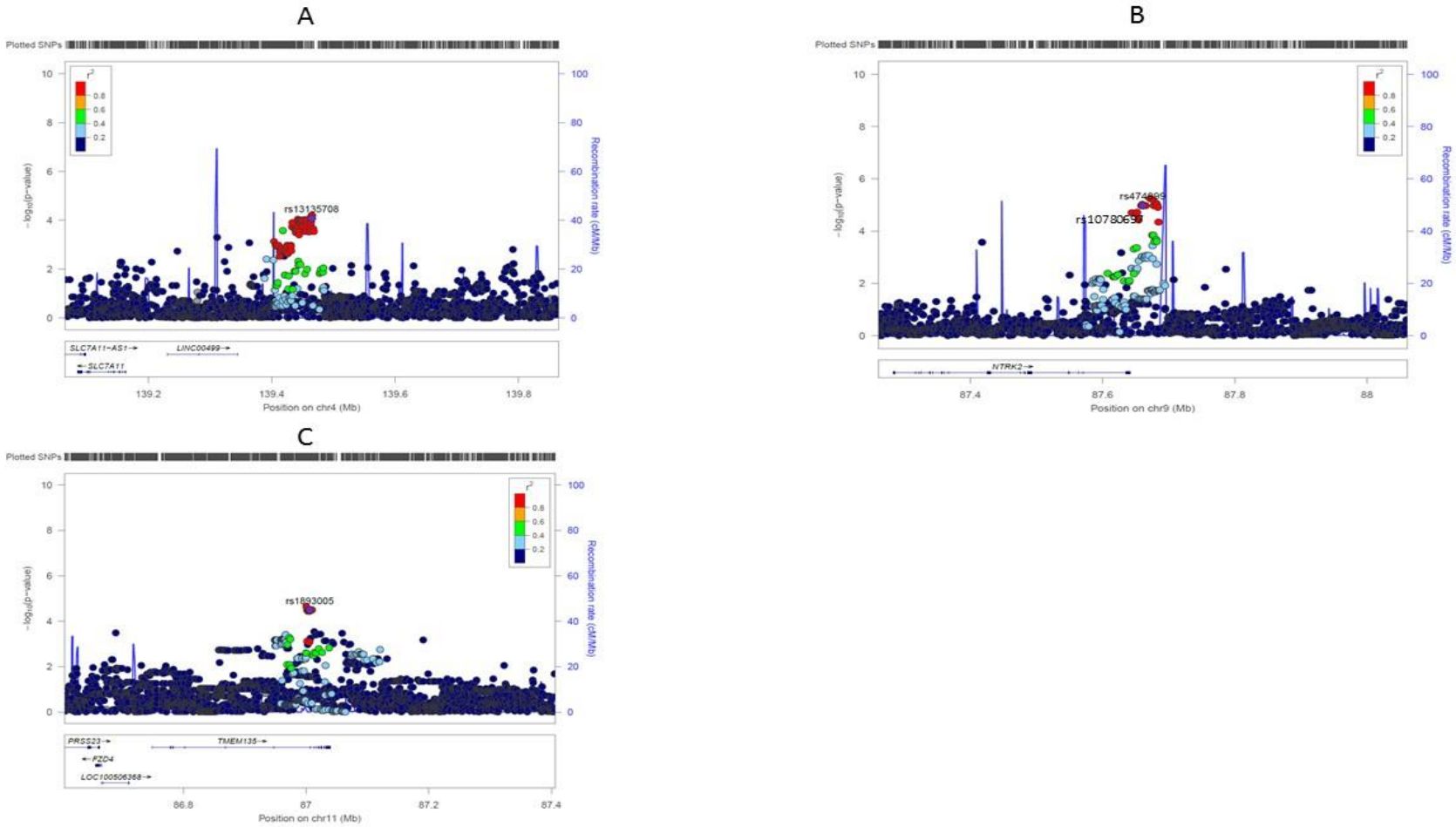


Figure 5.7 Locus Zoom plots for identified significant SNPs for bone stress injuries from chromosomes 4, 9 and 11. Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 3 in *NLGN1* gene; B, C) signals in the intragenic regions on chromosome 4.

5.3.4 Replication analysis

Imputed data was also utilised to attempt replication of the results from previously published studies, which used a candidate gene approach and identified genetic variants associated with the risk of bone stress injuries. *P*-values of eight significant SNPs located in six genes and described in the literature review were identified and summarised in Table 5.4. None of these previously identified SNPs reached statistical significance in this study. Locus oom plots (Figure 5.9 and Figure 5.10) for these eight SNPs showed very low significance profiles in the areas of the candidate genes. Currently, there are no other published GWAS investigating genetic markers of bone stress injuries, which limits opportunities in replication analyses.

Table 5.4 Summary of candidate SNPs from previous studies and their significance levels identified using imputed data.

SNP	Chr	BP	MAF	Gene	A1	OR	SE	*p*-value
rs1548456	7	93192036	0.28	*CALCR*	T	1.16	0.12	0.229
rs2051748	7	93199574	0.38	*CALCR*	G	0.76	0.12	0.0261
rs4328262	12	48285648	0.38	*VDR*	G	0.96	0.12	0.7136
rs1718119	12	121615103	0.32	*P2X7 (P2RX7)*	A	0.97	0.12	0.79
rs3751143	12	121622304	0.21	*P2X7 (P2RX7)*	C	0.87	0.16	0.373
rs1021188	13	43116133	0.22	*RANKL(TNFSF11)*	C	1.2	0.15	0.2285
rs1877632	17	41799590	0.27	*SOST*	A	1.01	0.13	0.961
rs3018362	18	60082093	0.38	*RANK (TNFRS11A)*	A	0.95	0.13	0.671

SNP – single nucleotide polymorphism, Chr – chromosome, BP – base pairs, MAF – minor allele frequency, A1 – effect allele, OR – odds ratio, SE – standard error.

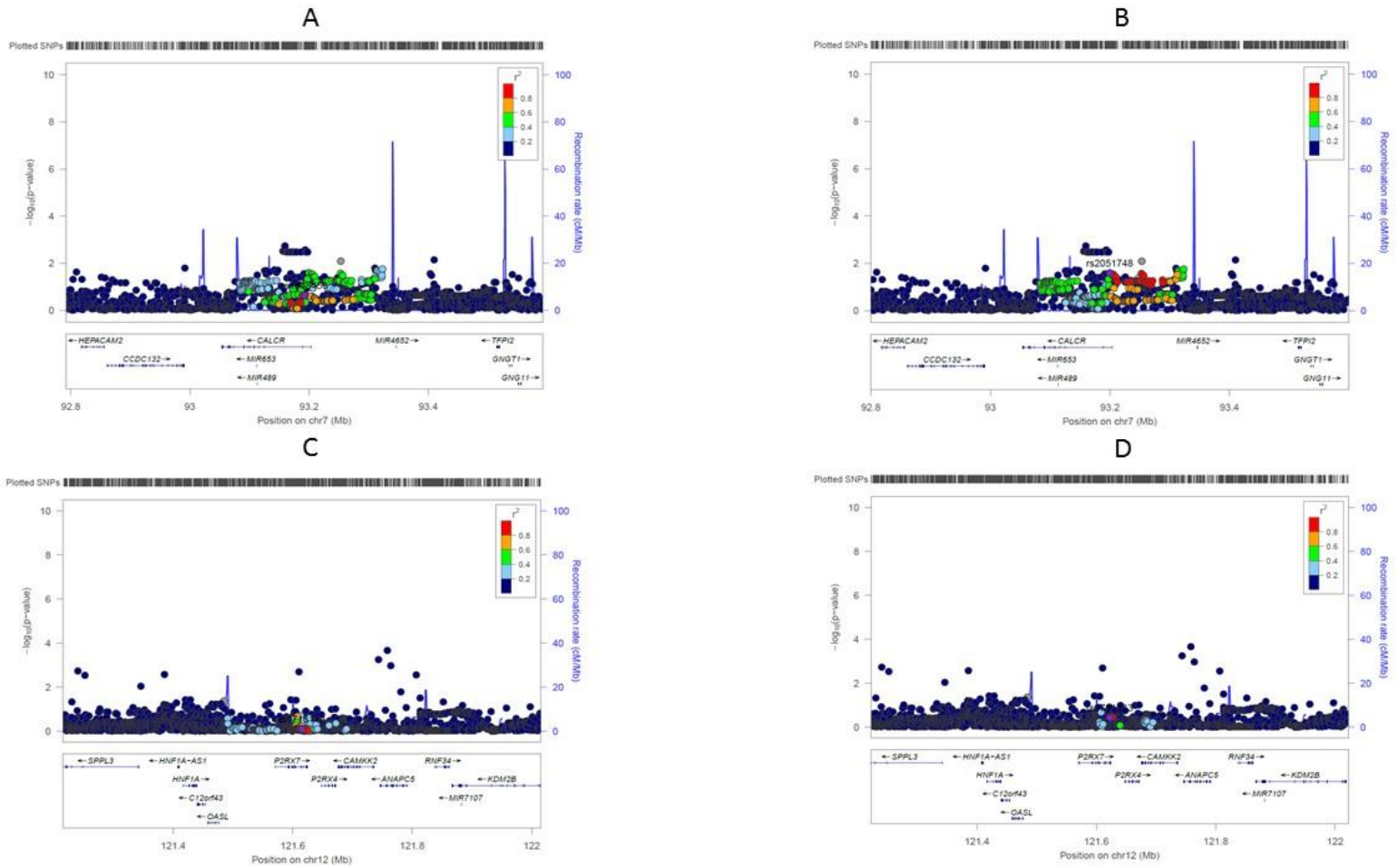


Figure 5.8 Locus Zoom plots of the SNPs from replication analysis for bone stress injuries from chromosomes 7 and 12. Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A, B) signals on chromosome 7 in *CALCR* gene; C,D) signals on chromosome 12 in *P2X7 (P2RX7)* gene.

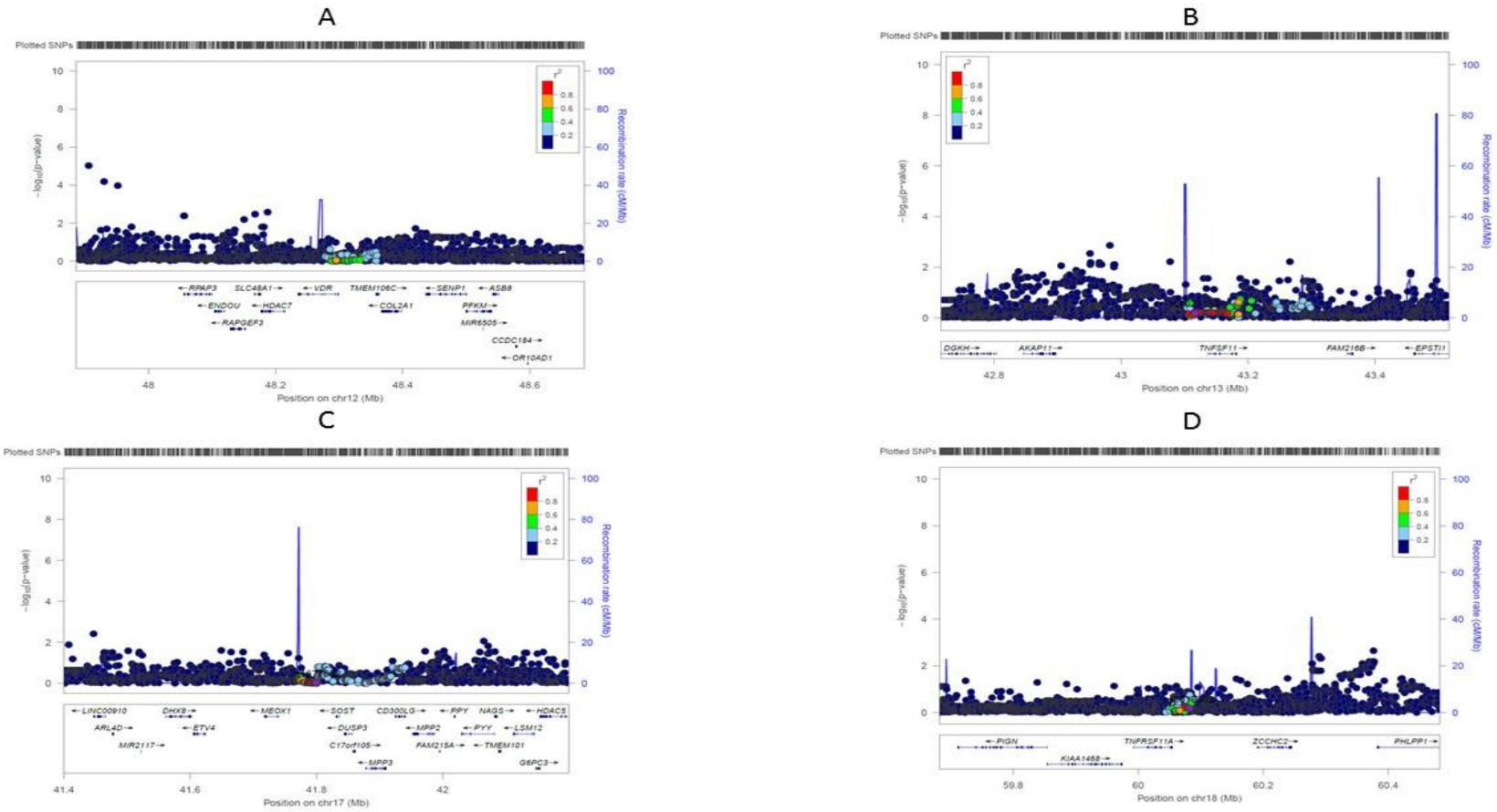


Figure 5.9 Locus Zoom plots of the SNPs from replication analysis for bone stress injuries from chromosomes 12, 13, 17 and 18.

Each SNP was plotted with 400kb flanking on both sides. Plotted SNPs are displayed in purple colour and named, surrounding SNPs are coloured based on their correlation (r^2) with genotyped SNP. An LD heat map is displayed in the r^2 legend in the top left corner of each plot. Genes where plotted SNPs are located, and surrounding genes are displayed in the bottom of each plot. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. A) signals on chromosome 12 in *VDR* gene; B) signals on chromosome 13 in *RANKL (TNFSF11)* gene; C) signals on chromosome 17 in *SOST* gene; C) signals on chromosome 18 in *RANK (TNFRS11A)* gene.

5.4 Discussion

Research of bone stress injuries in a study on twins and their high recurrence rates in athletes led to the investigation of genetic factors, which may contribute to the development of these injuries (Bennell et al., 1996b; Singer et al., 1990). However, the genetic contribution to bone stress injuries is represented by a limited number of studies and primarily based on findings from genetic studies on bone mineral density and disorders affecting bone health, such as osteoporosis and osteoarthritis. Thus, from the studies used a candidate gene approach, genes which protein products are involved in the metabolism of vitamin D and calcium, the Wnt signalling pathway and the RANK/RANKL/OPG signalling pathway were the main candidates.

The GWAS approach presented in the current study analysed 174 cases of bone stress injuries and 767 uninjured controls. All included recreational runners reported running at least 15 km per week and were between 18 and 50 years of age. Case-control analysis of genetic data did not identify any statistically significant SNPs reaching $p<5x10^{-8}$. Hence a greatly reduced threshold ($p<9.4x10^{-5}$) was employed to identify the top 20 most associated SNPs. Interpretation of these SNPs was supported with imputed data resulting in the identification of ten genes, which may be important in the development of bone stress injuries, specifically *LYPD6B*, *TMEM135, CSMD1, TEAD4, GSTA2, BACH2, RARS, NLGN1, PDZRN4* and *NTRK2*. These genes, included in the discussion, were assorted by their affiliation to a certain pathway of factors, which indirectly may influence processes in the bone. In addition, a study of gene expression in osteoblasts allowed us to compare expression levels of the selected genes and use them as supporting evidence for genes' potential role in bone stress injuries (Moriarity et al., 2015).

Two genes were linked to the Wnt pathway, which was previously identified as an important pathway in the regulation of bone metabolism (Krishnan, Bryant, & MacDougald, 2006). Transmembrane protein 135 (*TMEM135*) has been previously studied in relation to osteoporosis, and its expression has been identified in adipocytes and osteoclasts during their differentiation. *TMEM135* has also been identified as an important factor for osteoblastogenesis from human multipotent adipose tissue-derived stem cells (Scheideler et al., 2008). A meta-analysis of GWAS in 13 cohorts, which investigated genetic polymorphisms associated with heel bone properties such as broadband ultrasound attenuation (BUA), the velocity of sound (VOS), bone mineral density (BMD) showed that

the *TMEM135* gene was significantly associated with BUA and VOS (Moayyeri et al., 2014). In addition, it was shown that the structure of this transmembrane protein is homologous to a transmembrane region of frizzled-4, which is a component of the Wnt signalling pathway, previously identified as an important pathway in bone formation (Piters et al., 2008; Scheideler et al., 2008). However, the imputation results of this study showed that the SNP in *TMEM135* was not highly supported by other surrounding SNPs in high LD, which suggests that the function and the role of this gene require further research. The second gene LY6/PLAUR domain containing 6B (*LYPD6B*) results in a product which belongs to a lymphocyte antigen-6/urokinase-type plasminogen activator receptor superfamily with specific structural features, yet high variability by their function (Loughner et al., 2016). A review of the proteins in this family described Lypd6b, and its homolog Lypd6, as neuromodulators in mammals and detailed their involvement in the embryonic development (Vasilyeva, Loktyushov, Bychkov, Shenkarev, & Lyukmanova, 2017). The function of the protein encoded by *LYPD6B* is poorly studied, however, it was shown to be a target gene of the Wnt signalling pathway (Paramonov et al., 2017; hang et al., 2017). Thus, the discovered link between *LYPD6B* and the Wnt signalling pathway may be another supporting factor of this pathway's role in the bone stress injury susceptibility. However, the current limited knowledge of the function and regulation of this gene requires further investigation. Interestingly, the RNA expression study demonstrated that both of these genes were expressed in osteoblasts at the moderate level, which contributes to their potential for subsequent research.

Cytokines play an important role in osteoclast development and, therefore, in bone resorption (Manolagas, 1995). BTB domain and CNC homolog 2 (*BACH2*) is a gene that encodes a transcription factor, which is abundantly expressed in hematopoietic tissues, particularly in B cells and has been associated with lymphomas (Ichikawa et al., 2014; Sasaki et al., 2000). BACH2 also plays a crucial role in T-cell mediated immune responses and controls production of interleukin-4 (IL4) (Kuwahara et al., 2016), which in turn was identified as one of the inhibitors of osteoclast development (Manolagas, 2000). In addition, *BACH2* was identified as a locus associated with rheumatoid arthritis (Ruiz-Larra aga et al., 2016). This gene was also moderately expressed in osteoblasts (Moriarity et al., 2015). In the current study, imputation analysis demonstrated a high density of imputed supporting SNPs which were located in this gene. Therefore, although its function is not at present

directly linked to the functioning of bone cells, its involvement in cytokine regulation and immune response may justify the involvement of this gene is BSI susceptibility.

Both osteoclasts and osteoblasts are equipped with adrenergic and neuropeptide receptors, which are used for communication with the nervous system (Togari & Arai, 2008). The sympathetic nervous system is involved in the regulation of bone formation via controlling leptin production (Takeda et al., 2002). Several of the genes identified are linked to nervous system functioning. CUB and Sushi multiple domains (*CSMD1*) is a gene which is abundantly expressed in the central nervous system and epithelial tissues (Sun et al., 2001). PD domain-containing ring finger 4 (*PDZRN4*) is a gene which encodes a family member of Ligand of Numb Protein-X and therefore, is implicated in the cell fate determination through the inhibition of the Notch signalling pathway (Katoh & Katoh, 2004). Neuroligin 1 (NLGN1) is a splice site-specific ligand for β-neurexins and belongs to a family of neuronal cell surface proteins, which are involved in formation and remodelling of synaptic contacts in the central nervous system (Ichtchenko et al., 1995; Reissner, Klose, Fairless, & Missler, 2008). Neurotrophic receptor tyrosine kinase 2 (*NTRK2*) encodes membrane-bound kinase, which regulates the functioning of brain-derived neurotrophic factor (BDNF). BDNF is responsible for the key processes in neurological development such as neuronal birth, maturation, differentiation, migration and survival. This factor is essential for dendritic growth and synaptic plasticity (Huang & Reichardt, 2001). Additionally, genetic variation in the *NTRK2* gene was investigated in association with various psychiatric comorbidities. All these four genes were predominantly investigated in association with various neurological and psychiatric disorders (Baranzini et al., 2008; Lewis et al., 2010; Nakanishi et al., 2017; Pandya, Kutiyanawalla, & Pillai, 2013; Ray, Weickert, & Webster, 2014; Spalek et al., 2017; Torres et al., 2017; Voegeli et al., 2016; hang et al., 2015).

Glutathione S-transferase alpha 2 (*GSTA2*) is a gene associated with three of the top 20 SNPs from this study. The key function of its protein product is to detoxify carcinogens, therapeutic drugs, toxins and products of oxidative stress by conjugation with glutathione. This gene belongs to the alpha class of five transferases clustered on chromosome 6. This family is highly versatile and abundantly expressed in liver, kidneys and adrenal tissue (Seideg rd & Ekstr m, 1997). The function of this enzyme was extensively researched in association with cancer risk and utilisation of toxic drugs (McIlwain, Townsend, & Tew, 2006; hang, Mod n, & Mannervik, 2010). Imputation analysis showed that supporting SNPs of these three genotyped SNPs were located in neighbour genes encoding other GSTA family

members. However, these supporting SNPs were in low LD with genotyped SNPs. In addition, expression of *GSTA2* has not been observed in osteoblasts (Moriarity et al., 2015). All these findings suggest a contributory role of *GSTA2* into bone metabolism. In addition, the presence of three SNPs in this genomic area may be taken into consideration for further research.

The arginyl-tRNA synthetase (*RARS*) gene encodes for the cytoplasmic tRNA synthetase for arginine. This synthetase is part of a multienzyme complex, which plays a key role in translation and is expressed in all tissues and cells, including osteoblasts (Girjes, Hobson, Chen, & Lavin, 1995; Moriarity et al., 2015). Due to its essential presence in every cell, it is difficult to speculate on its specific role in bone cells and the contribution of this gene variation to bone stress injury development.

The remaining gene, TEA domain transcription factor 4 (*TEAD4*), encodes a transcriptional enhancer factor (TEF) and is also known as TEF3. This gene is typically expressed in skeletal muscle cells, promoting myoblast differentiation and epithelial-mesenchymal transition (Benhaddou et al., 2012). This gene has no known role in relation to bone formation and to date has predominantly been studied in embryonic cells and in relation to a role in cancer (Shi et al., 2017; Skottman et al., 2005).

In conclusion, the most interesting finding was an association of *TMEM135* variation with bone stress injuries. This gene encodes a transmembrane protein which regulates the growth of osteoblasts and is identified as one of the genetic markers associated with heel bone properties (Moayyeri et al., 2014). *TMEM135* and *LYPD6B* both were functionally associated with the Wnt signalling pathway, which allowed us to flag them as genes of interest for further research. In addition, a review of gene expression analysis (via RNA sequencing) in normal human osteoblast cells showed that some of the previously discussed genes had relatively high (*RARS*) or moderate (*TMEM135, TEAD4* and *LYPD6B*) expression levels, whilst expression of *GSTA2* and *PDZRN4* were undetected (Moriarity et al., 2015). Among the reviewed genes, four were predominantly expressed in the neuronal system and responsible for neurological development, the functioning of the central nervous system and synaptic plasticity. Several mouse models with leptin deficiency demonstrated the existence of the regulation of bone formation by the central nervous system (Elefteriou, 2008). Another study showed co-expression of *NTRK2* and *BDNF* in the active phase of osteoblasts in a rat model, which suggested their regulatory role in bone remodelling (Yamashiro, Fukunaga, Yamashita, Kobashi, & Takano-Yamamoto, 2001). The nervous

system plays a major role in bone formation and remodelling, which supports the presence of genetic polymorphisms in *PZDRN4 CSMD1, NTRK2* and *NLGN1* associated with bone stress injuries. This provides additional evidence of the complex interaction between bone and nervous system.

Research of multifactorial conditions is complicated by the interplay of genetic and environmental factors, their variation and their unclear proportions of contribution to the cause of the condition. Additionally, the effect of genetic factors of multifactorial conditions such as overuse injuries is usually very low and varies between 1.2 and 1.5 OR (Bodmer & Bonilla, 2008). Therefore, the risk of the condition between carriers and non-carriers of certain genetic variants may differ marginally. This issue may be compensated by a sample size large enough (typically in the thousands) to identify these low effect polymorphisms. However, the recruitment of a cohort of sufficient size for GWAS can be very challenging, as previously discussed (Section 4.4.3). In addition, the review of the functions and expression location of the selected genes suggest that the majority of genes in this study were not associated with the processes responsible for bone formation and metabolism. This could be explained by the high chance of false-positive results due to the decreased *p*-value threshold. In addition, the attempted replication analysis demonstrated that none of the previously described genetic variants reached the significance level. Unfortunately, the genetics of bone stress injuries is currently poorly understood and requires additional data collection and recruitment of physically active people experiencing such injuries.

Although the identified genetic variants may have a weak effect and low predictive value, exploration of their function in molecular pathways may potentially reveal previously unknown processes or biomarkers that contribute to the development of an injury. The GWAS approach may be an efficient and unbiased tool to explore genetic predisposition to bone stress injuries, yet, in the current study, will require a significant increase of the sample size either via additional recruitment or through collaboration with similar research projects and data merging.

6. Conclusion – Key findings and future directions

6.1 Introduction

As the original project aimed to recruit approximately 10,000 Australian recreational runners, the initial data and sample collection methods were designed to be minimally invasive, using online survey software and remote collection of saliva samples. The survey was designed to collect detailed health and injury data from Australian recreational runners that allowed us to identify runners eligible for the genetic part of the project (Domaschenz et al., 2015). A GWAS approach was selected as an unbiased method for identification of genetic variants associated with the studied injury traits. This method allowed us to analyse a large sample size in a short period and utilise programming software with predetermined workflow for the subsequent analyses of the high-volume genetic data. As a large number of recruited participants was a critical factor in reaching the significance level in the genetic analysis, the recruitment phase was the longest part of the project, taking 25 months. Versatile recruitment strategies were employed throughout this period; including the online advertisement, paid Facebook promotion and attendance of running events (Manzanero et al., 2018). As a result of the recruitment efforts, data from 4,720 responses were utilised for epidemiological data analysis and 1,099 DNA samples were used for investigating genetic variants associated with overuse running-related injuries. Statistical analyses of the collected epidemiological data provided insights to the training and lifestyle habits, and injury history of Australian recreational runners. In addition, risk factors of the most common running-related injuries were identified and discussed. The genetic analysis allowed us to identify genetic variants associated with Achilles tendon injuries and bone stress injuries at the suggested significance level of $p<1x10^{-3.9}$ and $p<9.4x10^{-5}$, respectively, and then compare these associations previously published associations.

6.2 Conclusions from epidemiological data analyses

Initially, the epidemiological data analysis aimed to describe physical, training and lifestyle characteristics of the Australian recreational runners (Chapter 2). The recruited recreational runners displayed lifestyle and training characteristics, which were associated with a range of beneficial health outcomes. The recreational runners included in the study reported running at least 15 km per week, and 70 of the respondents running more than 20 km per week within at least two running sessions. Moreover, three-quarters of the respondents also reported participating in sports other than running. Thus, we concluded that recreational runners in the studied cohort met physical activity guidelines suggested by the WHO (World Health Organisation, 2010). Interestingly, reports of smoking, excessive weight and hypertension were uncommon for the participants. These findings suggested that recreational runners predominantly avoided the key contributors to the burden of disease (Australian Institute of Health and Welfare, 2016b). The data also demonstrated that the commencement of a regular running program was associated with a significant weight loss irrespective of biological sex, participation in other sports and injury history. This association between recreational running and healthy characteristics may contribute to the promotion of recreational running as a regular exercise for its beneficial effects such as improvement in aerobic capacity, weight loss and maintenance of overall healthy lifestyle. In addition, the data analysis demonstrated differences in training characteristics of male and female runners, as male runners were more likely to run longer weekly distances at higher frequencies of the weekly running sessions than female runners. These differences in the running patterns, together with the previously reported different motivation reasons for exercise between men and women (Australian Sports Commission, 2016; Lauderdale et al., 2015) may assist in targeted marketing of running among different groups of the population.

The second study (Chapter 3) demonstrated injury occurrence rates and common types of injuries typical for Australian recreational runners. Thus, over 50 of the recruited recreational runners reported running-related injuries, which had occurred in the past two years. Statistical analyses of the collected injury data identified the two most common reported injuries among Australian recreational runners as Achilles tendon injuries and bone stress injuries. Achilles tendon injuries were more typical for male runners, whereas the rates of the reported bone stress injuries were similar for male and female runners. Multiple logistic regression analyses aimed to identify factors associated with the occurrence of running-related injuries in general, and Achilles tendon injuries and bone stress injuries

independently. The multiple logistic regression analysis identified factors associated with a general risk of running-related injuries, such as male sex, younger age, and stretching in relation to a running session. Factors associated with Achilles tendon injuries comprised male sex, age over 35 years, faster race pace of 4min/km and stretching in relation to a running session. The identified factors associated with the development of bone stress injuries were age under 35 years, high BMI of over 30 kg/m^2 categorised as obese, the weekly running distance of longer than 40 km and stretching in relation to a running session. These findings demonstrated the diversity of the factors which should be considered in the training programs aiming to prevent injuries. Identified training-related factors associated with the development of these types of injuries may contribute to the further research of running-related injuries. In addition, better knowledge of these factors may assist with the development of training programs and reduce injury rates among recreational runners.

6.3 Conclusions from GWAS of Achilles tendon injuries and bone stress injuries

In this project, two GWAS were conducted in order to identify genetic variants associated with Achilles tendon injuries and bone stress injuries (Chapter 4 and Chapter 5, respectively). The statistical case-control analyses of the genomic data were conducted independently for each type of injury. The analysis of the genetic variants associated with Achilles tendon injuries was conducted on 171 cases and 767 uninjured controls. Although this limited sample size did not allow us to reach the standard GWAS significance threshold $p<5x10^{-8}$, an exploratory examination of the 20 most significant genetic variants was performed, further supported by SNP imputation and protein function. Several of these genetic variants were located in the genes *TCF7L1*, *DOCK4* and *TLE1,* whose functions are associated with the Wnt signalling pathway. The Wnt pathway regulates transcriptional pathways in stem cell activation, cell proliferation and differentiation. The Wnt pathway also targets MMPs, which regulate the homeostasis of the extracellular matrix in the tendon (Clevers, 2006). Our findings suggest that the Wnt signalling may be involved in the regulation of homeostasis of the tendon matrix and therefore further research of its function and the potential link to the pathological processes occurring in the tendon may assist in better understanding of the tendinopathy. In addition, the imputation of genetic variants allowed us to compare our finding to the results of a previously published GWAS, which investigated genetic variants associated with Achilles tendinopathy in more than 100,000 hospital patients (Kim et al., 2017). This study also did not reach the stringent $p<5x10^{-8}$ significance threshold. An attempted cross-replication of this study's results and the GWAS results performed by Kim et al. (2017) failed. Furthermore, this study also failed to replicate associations identified in previous candidate gene studies (Section 4.3.6.3). These analyses demonstrated that none of the genes involved in the collagen structure (*COL5A1, COL11A1*), apoptosis (*CASP3, CASP8*) and inflammation (*IL1B*, *IL1RN, IL6*) reached the suggestive significance level of $p<1x10^{-3.9}$. This study's failed replication of the previous candidate gene studies' results is supported by Kim et al. (2017), who also failed to replicate these candidate gene associations (Kim et al., 2017).

The GWAS of genetic variants associated with bone stress injuries was conducted on 174 cases and 767 uninjured controls. This sample size made the study underpowered; therefore, the top 20 significant signals were investigated in a highly exploratory manner, following a *p*-value threshold reduction from $p<5x10^{-8}$ to $p<9.4x10^{-5}$. The functions of the genes in the list of top 20 most significant variants were highly variable and only one of the

explored genes – *TMEM135,* potentially had a function relevant to the development of bone stress injuries. This gene was characterised as an important factor for osteoblastogenesis and was previously investigated in association with another condition affecting bone health – osteoporosis (Scheideler et al., 2008). Another gene *LYPD6B* encodes a product, which is involved in the embryonic development, however, is poorly studied. Importantly, *TMEM135* and *LYPD6B* genes were both associated with the Wnt signalling pathway. Although the Wnt signalling pathway's functions are versatile, its malfunction was previously associated with osteopenia or osteoporosis. Therefore, the described association suggested further research towards a better understanding of the Wnt pathway in bone formation. Genetic markers of bone stress injuries were poorly studied, however, several candidate gene approach studies suggested that genetic variants of *VDR, CALCR, P2X7, RANKL, RANK, SOST* genes were associated with the bone stress injuries (Styrkarsdottir et al., 2008; Varley et al., 2016; Yanovich et al., 2012; muda et al., 2011). Therefore, we attempted to replicate these results using GWAS data. The genetic variants of the genes as mentioned above did not reach the suggested significance level of $p<9.4x10^{-5}$. These results suggest that the genetics of bone stress injuries requires more research, which may enable replication and comparison of results from different studies.

6.4 Key findings

1. Australian recreational runners demonstrated healthy lifestyle habits, including meeting physical activity guidelines recommendations (Section 2.4.1).

2. Commencement of a running program was associated with a significant weight loss in the novice recreational runners (Table 2.9).

3. Male and female runners had different characteristics in running programs, and therefore, the appropriate promotion of physical activity may depend on gender (Table 2.6).

4. Achilles tendon injuries and bone stress injuries were two most commonly reported running-related injuries in the cohort of Australian recreational runners (Figure 3.1).

5. Male recreational runners were more likely to report running-related injuries than female runners (Table 3.4).

6. Faster running race pace, older age and stretching were the factors associated with the reporting of Achilles tendon injuries in the population of Australian recreational runners (Table 3.7).

7. Longer running distance, younger age, obesity and stretching were associated with the reporting of bone stress injuries in the population of Australian recreational runners (Table 3.10).

8. Genetic variants in genes linked to the Wnt signalling pathway (*TCF7L1*, *DOCK4* and *TLE1*) may contribute to the occurrence of Achilles tendon injuries (Figure 4.4, Table 4.4). This observation is highly exploratory and requires further research to substantiate.

9. The study was unable to replicate genetic results from the previous candidate gene approach studies and one GWAS, which may suggest that associations of the certain genetic variants are likely to be study-specific (Table 4.5, Table 4.6).

10. The study was unable to determine the genetic polymorphisms associated with bone stress injuries. Only two genes (*TMEM135* and *LYPD6B*) suggested a link between the Wnt signalling pathway and the bone stress injury occurrence (Figure 5.2, Table

5.3). Given the impact of bone stress injuries on participation in physical activity, this requires further research to substantiate.

11. The study failed to replicate results from the candidate gene approach studies, which suggested that several genes were associated with the development of bone stress injuries (Table 5.4).

12. Although three polymorphisms were a putative replication: rs1110495 (16:51914974, *p* 0.0215), rs12722 (*COL5A1* gene, *p* 0.0194) and rs2051748 (*CALCR* gene, *p* 0.0261), these results should be interpreted with caution due to imputed nature of the data (Tables 4.5, 4.6 and 5.4).

13. The genetic data collected and analysed in this study may assist in further research of the genetic variation associated with running-related injuries through replication, data pooling and meta-analysis.

6.5 Limitations

This cross-sectional study was based on large-scale recruitment of Australian recreational runners. The recruitment campaign lasted for 25 months, which was the most prolonged phase of the project. The recruitment campaign employed multiple strategies as they aimed to collect data from thousands of recreational runners. The study was promoted as the 'AIS Running Injury Study', which was described as a study of running-related injuries in Australian recreational runners, and research of genetic variants associated with Achilles tendon injuries and bone stress injuries. Thus, recreational runners provided their personal, health and injury data, which were analysed in the epidemiological part of the study (Chapters 2 and 3). The collected survey data was of a self-reported retrospective nature, and previous studies demonstrated limitations in collecting sporting injury data retrospectively (Gabbe et al., 2003; Kolt & Kirkby, 1999). However, the utilised questionnaire had been demonstrated to provide reliable data for this project (Domaschenz et al., 2015). The final dataset comprised over 4,700 completed responses, which was the largest known dataset of health and injury data collected from Australian recreational runners. As previously discussed (Section 2.5.1), there were approximately 2.8 million recreational runners in Australia during the recruitment phase (Australian Sports Commission, 2016). Although this study was vastly promoted by the Australian Institute of Sport – a well-known sporting institution in Australia, the recruitment of initially planned 10,000 recreational runners was not achieved. Whilst several recruitment strategies across Australia were employed, including social and online media and attending running events in other states the highest participation rate (7x) per 100,000 was observed in the ACT, where the AIS is located. This clearly highlights the importance of multisite recruitment for future projects to further maximise participant recruitment. The genetic arm of the study aimed to investigate genetic variants associated with Achilles tendon injuries and bone stress injuries, which were more likely to occur due to running overload, stringent selection criteria were applied to the studied cohort. Only runners younger than 50 years of age, who avoided smoking, did not experience any fractures or chemotherapy, and were not diagnosed with any chronic conditions which affect the musculoskeletal system, were eligible to participate in the genetic analysis. Therefore, to reach the sample numbers of at least 800 cases and 800 controls required for GWAS was also a significant challenge. In addition, as respondents were requested to consent to provide genetic material for the following GWAS in the survey, we identified that younger runners were less likely to

agree to participate in the genetic component of the project (Manzanero et al., 2018). Interestingly, a similar trend was observed in another independent study (McQuillan & Porter, 2011). With the reluctance to provide genetic material for research among younger adults, possibly due to lower levels of trust in research and greater privacy concerns.

Performing GWAS for complex diseases is challenging due to the multiple genetic and environmental factors involved. The GWAS performed in this presented study, required careful planning and design, including the estimation of disease prevalence and heritability, choice of genotyping chip, required a sample size and calculated study power (Spencer, Su, Donnelly, & Marchini, 2009; ondervan & Cardon, 2007). At the commencement of this study there was no other large-scale genetic variant analyses of Achilles tendon injuries and bone stress injuries. Therefore, this study would be the first of its kind to investigate high-density genetic predisposition in these two injury types. This project was an initiative of the CRN-AESS, and similar to other genetic-based studies, there were challenges with recruiting sufficiently large numbers of participants. As previously demonstrated (Section 1.8.3), a large sample size would allow us to investigate SNPs with relatively low MAFs and effect size. The OR of a single associated SNP is typically small, with the majority of SNPs associated with multifactorial conditions varying between OR of 1.2 to 1.5 (Bodmer & Bonilla, 2008). As previously shown (Section 1.8.3) our initial power calculations based on recruitment of 800 cases, relied upon an OR 1.5 or greater. Therefore, this study would only be able to identify SNPs if their OR was at the upper end of this range. The probability of identifying SNPs with a lower OR was therefore substantially reduced using the standard, stringent statistical thresholds. Their identification relied on the recruitment of tens of thousands of participants, which was well beyond the scope and timeframe of this study. Therefore, these types of studies need to employ recruitment strategies that use multiple locations and international collaborations to have the greatest impact.

Another limitation may be due to the selection criteria used in this study. These relatively strict selection criteria were employed to control for confounding factors and therefore focus on the examination of genetic variants associated with these two injuries in physically active, healthy individuals. However, the application of this stricter selection criteria significantly reduced the number of eligible study participants from our original survey population. Using more relaxed selection criteria would have more likely enabled

the recruitment of the target sample size of 2,400 participants (800 in each of three analysed groups). However, this larger recruited population would be more heterogeneous, and identification of statistically significant genetic variants may be confounded by other uncontrolled factors.

As GWAS have become a standard research approach, their value increases when other research groups recruit similar cohorts, allowing data to be pooled, or results replicated. For example, GWAS by Kim et al. also failed to identify statistically significant associations with Achilles tendon injuries employing 5,000 cases and over 100,000 controls (Kim et al., 2017). In addition, this large study also failed to replicate results from the previous candidate gene studies. However, as acknowledged by the authors of this study, their study participants were not controlled for physical activity levels, which may have had a strong confounding effect, diluting out any associations. Collectively this demonstrates a delicate balance between applying strict selection criteria and therefore limiting sample size, versus recruiting a large heterogeneous cohort, and the analysis being affected by multiple confounding factors. This is further evidenced by a GWAS that identified *FTO* as an obesity susceptibility gene, through the analysis of 2,000 cases with type 2 diabetes and 3,000 controls across five centres in the UK (Frayling et al., 2007). At first, a cluster of SNPs in the *FTO* gene was found to be associated with type 2 diabetes (p $5*10^{-8}$). Several replication GWAS confirmed that genetic variants in the *FTO* gene were associated with susceptibility to obesity (Dina et al., 2007; Scuteri et al., 2007). However, further replication of these results and adjusting for BMI as a confounder abolished this association (Loos & Yeo, 2014). This example demonstrates that such factors as sample size, multisite research, correctly controlling for confounders and replication of findings together would advance GWAS research and allow researchers to identify robust genetic associations.

The present study was focused on exercise-related injuries in healthy individuals, which other than the Kim et al. (2017) study, restricted the opportunity to replicate results from other GWAS. In addition to attempted replications, the UK Biobank health resource, which provides access to GWAS data across multiple traits, was explored as a source of similar studies (UK Biobank, 2019). However, this resource did not contain any studies (other than Kim et al. (2107)) of genetic associations of tendon or bone injuries with exercise-related phenotype, which was the key component of this study design.

As this research project aimed to identify factors associated with the development of running-related injuries, the literature review and hypotheses of this study relied on the previous finding from the exercise and sport science. This field of science has developed in recent years and comprises several disciplines, such as physiology, biomechanics, nutrition, sports medicine (Halperin, Vigotsky, Foster, & Pyne, 2018). In addition to the previously discussed challenges of GWAS research, exercise and sports science research also has its methodological issues. A recent review by Halperin et al. (2018) outlines several of these problems and their possible solutions to improve the quality of results in sport science (Halperin et al., 2018). Specifically, that exercise and sport science research would benefit from more validation and replication studies, access and sharing of raw data and collaborations (Halperin et al., 2018).

These indicated issues suggest that this interdisciplinary project had several limitations due to the existing challenges in exercise science and GWAS. In summary, we again acknowledge that the genetic associations and interpretations in this study are currently of an exploratory nature. As such, they should be interpreted with caution, due to the limited sample size and reduced statistical thresholds employed.

6.6 Future directions

Research into the genetic factors of multifactorial conditions is complex due to the interaction of genetic and environmental factors, which may contribute to the development of these conditions to varying degrees. However, typically the effect of genetic variants to the development of multifactorial conditions is rather low (Bodmer & Bonilla, 2008). With the development of high throughput genetic technologies, in particular the GWAS approach, genetic research has become more accessible and cost-effective, allowing for the investigation of a wider range of multifactorial conditions. However, these studies require consistency in the collected data and distinct phenotypic characteristics, which may restrict the sample size. The challenge of recruiting large numbers of participants may be resolved by collaborative efforts across multiple research teams and consortia. Similarly, the sharing of phenotypic and genetic data, and merging datasets with similar characteristics would allow researchers to increase final sample size and perform data meta-analysis.

As this project was a part of a large collaborative network, phenotypic and GWAS data from this study was made available to collaborating research teams and will be accessible to more research groups in the future. The GWAS data from this study has already been utilised as part of the control group for another project investigating necroptotic cell death (Hildebrand, 2019). This was a large collaborative project that utilised cell biology and GWAS to identify polymorphisms in the *MLKL* gene, and assist in the understanding of the underlying processes of paediatric autoinflammatory disease (Hildebrand, 2019). Furthermore, sharing GWAS data may contribute to the further research of Achilles tendon injuries and bone stress injuries by adding data value to the meta-analyses.

Conversely, this study significantly benefited from the freely available GWAS data generated by Kim et al (Kim et al., 2017). Access to these data enabled the ability to perform replication of our study findings in a separate study cohort. Whilst replication of our data could not be achieved and vice versa, this emphasised another important issue regarding the genetic research of multifactorial conditions, namely the independent reproducibility of results. Therefore, the current explosion of direct-to-consumer genetic testing for exercise, nutrition and sports, including injury risk, should be considered with caution. These tests rely on the results from the candidate gene approach studies, which, for instance, identified associations between polymorphisms in *COL5A1*, *TNC* and *MMP3* genes with the risk of Achilles tendon injuries. However, none of these independent replication studies could provide any evidence that these polymorphisms were predictive of the injury risk, and would

have similar effects across various ethnicities and both sexes. Extreme caution must therefore be taken when interpreting results from direct-to-consumer tests which claim to provide advice on injury risk (Vlahovich et al., 2017). A recent joint statement from the AIS, the International Federation of Sports Medicine and the Athlome Consortium concludes that there is no place for these tests in predicting injury risk. Similarly the AIS has taken a position on the use of these commercial tests in athletes, advising against doing so (Vlahovich, Fricker, Brown, & Hughes, 2016). In the future, when more research can provide satisfactory reproducible scientific evidence of genetic variants associated with the injury risk, these tests may benefit injury prevention, but currently the predictive power of these tests is negligible. Therefore, the genetic results, demonstrated in this study, may be utilised for the future attempts to compare genetic associations and reproduce these results. In order to gather more genetic data and explore the contribution of genetics to injury risk, effective collaboration and communication between research groups should be the key part of this type of research. This could improve the outcomes of the replication analyses, lead to better quality results produced by individual studies and collaborations, and provide more scientific evidence in this topic.

This study was the first project involving GWAS research carried out by the AIS in collaboration within the CRN-AESS. This may provide the opportunity for the further genetic research in sports science and injury prevention. In conclusion, this project contributed to the understanding of running habits and injuries in Australian recreational runners and initiated GWAS analysis of Achilles tendon injuries and bone stress injuries in physically active, healthy individuals. Furthermore, these GWAS data are available for the further research of genetic variants associated with running-related injuries and other genetic research.

7. References

Abate, M., Oliva, F., Schiavone, C., & Salini, V. (2012). Achilles tendinopathy in amateur runners: role of adiposity (Tendinopathies and obesity). *Muscles, Ligaments And Tendons Journal, 2*(1), 44-48.

Abate, M., Salini, V., & Schiavone, C. (2015). Achilles tendinopathy in elderly subjects with type II diabetes: the role of sport activities. *Aging Clin Exp Res*. doi:10.1007/s40520-015-0391-7

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics, 3*, 103-107.

Abrahams, Y., Laguette, M. J., Prince, S., & Collins, M. (2013). Polymorphisms within the COL5A1 3'-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Ann Hum Genet, 77*(3), 204-214. doi:10.1111/ahg.12013

Armamento-Villareal, R., Napoli, N., Diemer, K., Watkins, M., Civitelli, R., Teitelbaum, S., & Novack, D. (2009). Bone turnover in bone biopsies of patients with low-energy cortical fractures receiving bisphosphonates: a case series. *Calcified Tissue International, 85*(1), 37-44.

Australian Bureau of Statistics. (2015). *Participation in Sport and Physical Recreation, Australia, 2013-14*.

Australian Bureau of Statistics. (2016a). *National Health Survey. First results. Australia 2014-15*.

Australian Bureau of Statistics. (2016b). Population by Age and Sex, Regions of Australia, 2016 Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3235.0Main+Features12016?OpenDocument

Australian Institute of Health and Welfare. (2016a). *Australia's health 2016*. Canberra Retrieved from http://www.aihw.gov.au/publication-detail/?id=60129555544.

Australian Institute of Health and Welfare. (2016b). *Australian Burden of Disease Study: impact and causes of illness and death in Australia 2011*. Canberra: AIHW.

Australian Institute of Health and Welfare. (2017). High blood pressure. Retrieved from http://www.aihw.gov.au/risk-factors/high-blood-pressure/

Australian Institute of Health and Welfare. (2018, 19/03/2019). The Health of Australia's Females. Retrieved from https://www.aihw.gov.au/reports/men-women/female-health/contents/who-are

Australian Sports Commission. (2016). *AusPlay Participation data for the sport sector*.

Bakrin, I. H., Hussain, F. A., & Tuan, S. E. (2016). Transducer-like enhancer of split 1 (TLE1) expression as a diagnostic immunohistochemical marker for synovial sarcoma and its association with morphological features. *The Malaysian journal of pathology, 38*(2), 117-122.

Baldwin Jr, A. S. (2001). Series introduction: the transcription factor NF-κB and human disease. *Journal of Clinical Investigation, 107*(1), 3.

Baranzini, S. E., Wang, J., Gibson, R. A., Galwey, N., Naegelin, Y., Barkhof, F., . . . Johnson, M. R. (2008). Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Human molecular genetics, 18*(4), 767-778.

Barfod, K. (2014). *Achilles tendon rupture; Assessment of nonoperative treatment* (Vol. 61).

Barrack, M. T., Gibbs, J. C., De Souza, M. J., Williams, N. I., Nichols, J. F., Rauh, M. J., & Nattiv, A. (2014). Higher Incidence of Bone Stress Injuries With Increasing Female Athlete Triad–Related Risk Factors: A Prospective Multisite Study of Exercising Girls and Women. *The American journal of sports medicine, 42*(4), 949-958.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2004). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*(2), 263-265.

Barton, C., Bonanno, D., Carr, J., Neal, B., Malliaras, P., Franklyn-Miller, A., & Menz, H. (2016). Running retraining to treat lower limb injuries: a mixed-methods study of current evidence synthesised with expert opinion. *Br J Sports Med, 50*(9), 513-526.

BC, P. (2015). BC Platforms software. Retrieved from http://bcplatforms.com/

Beck, B. R., Bergman, A. G., Miner, M., Arendt, E. A., Klevansky, A. B., Matheson, G. O., . . . Marcus, R. (2012). Tibial stress injury: relationship of radiographic, nuclear medicine bone scanning, MR imaging, and CT severity grades to clinical severity and time to healing. *Radiology, 263*(3), 811-818.

Beck, B. R., Rudolph, K., Matheson, G. O., Bergman, A. G., & Norling, T. L. (2015). Risk Factors for Tibial Stress Injuries: A Case–Control Study. *Clinical Journal of Sport Medicine, 25*(3), 230-236.

Benhaddou, A., Keime, C., Ye, T., Morlon, A., Michel, I., Jost, B., . . . Davidson, I. (2012). Transcription factor TEAD4 regulates expression of myogenin and the unfolded protein response genes during C2C12 cell differentiation. *Cell death and differentiation, 19*(2), 220.

Benjamin, M., Kaiser, E., & Milz, S. (2008). Structure-function relationships in tendons: a review. *Journal of anatomy, 212*(3), 211-228.

Bennell, K., Matheson, G., Meeuwisse, W., & Brukner, P. (1999). Risk factors for stress fractures. *Sports medicine, 28*(2), 91-122.

Bennell, K. L., Brukner, P. D., & Malcolm, S. A. (1996a). Effect of altered reproductive function and lowered testosterone levels on bone density in male endurance athletes. *British Journal of Sports Medicine, 30*(3), 205-208.

Bennell, K. L., Malcolm, S. A., Thomas, S. A., Reid, S. J., Brukner, P. D., Ebeling, P. R., & Wark, J. D. (1996b). Risk Factors for Stress Fractures in Track and Field Athletes A Twelve-Month Prospective Study. *The American journal of sports medicine, 24*(6), 810-818.

Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature genetics, 40*(6), 695.

Bouchard, C., Leon, A. S., Rao, D., Skinner, J. S., Wilmore, J. H., & Gagnon, J. (1995). The HERITAGE family study. Aims, design, and measurement protocol. *Medicine and science in sports and exercise, 27*(5), 721-729.

Brown, K. L., Seale, K. B., El Khoury, L. Y., Posthumus, M., Ribbans, W. J., Raleigh, S. M., . . . September, A. V. (2016). Polymorphisms within the COL5A1 gene and regulators of the extracellular matrix modify the risk of Achilles tendon pathology in a British case-control study. *Journal of Sports Sciences*, 1-9.

Buist, I., Bredeweg, S. W., Lemmink, K. A., Van Mechelen, W., & Diercks, R. L. (2010). Predictors of running-related injuries in novice runners enrolled in a systematic training program. *The American journal of sports medicine, 38*(2), 273-280.

Buscarlet, M., Hermann, R., Lo, R., Tang, Y., Joachim, K., & Stifani, S. (2009). Cofactor-activated phosphorylation is required for inhibition of cortical neuron differentiation by Groucho/TLE1. *PLoS One, 4*(12), e8107.

Cabello-Verrugio, C., Acuña, M. J., Morales, M. G., Becerra, A., Simon, F., & Brandan, E. (2011). Fibrotic response induced by angiotensin-II requires NAD (P) H oxidase-induced reactive oxygen species (ROS) in skeletal muscle cells. *Biochemical and biophysical research communications, 410*(3), 665-670.

Cavalli-Sforza, L. L. (2005). The human genome diversity project: past, present and future. *Nature Reviews Genetics, 6*(4), 333.

Chakravarty, E. F., Hubert, H. B., Lingala, V. B., & Fries, J. F. (2008). Reduced disability and mortality among aging runners: a 21-year longitudinal study. *Archives of internal medicine, 168*(15), 1638-1646.

Clark, I. M., Swingler, T. E., Sampieri, C. L., & Edwards, D. R. (2008). The regulation of matrix metalloproteinases and their inhibitors. *The international journal of biochemistry & cell biology, 40*(6), 1362-1378.

Clevers, H. (2006). Wnt/β-catenin signaling in development and disease. *Cell, 127*(3), 469-480.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second Edition ed.): Lawrence Erlbaum Associates, Inc.

Collins, M. (2010). Genetic risk factors for soft-tissue injuries 101: a practical summary to help clinicians understand the role of genetics and 'personalised medicine'. In: British Association of Sport and Excercise Medicine.

Cook, J. L., Khan, K. M., & Purdam, C. (2002). Achilles tendinopathy. *Manual therapy, 7*(3), 121-130.

Cook, J. L., & Purdam, C. R. (2009). Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *British Journal of Sports Medicine, 43*(6), 409-416.

Cook, J. L., Rio, E., Purdam, C. R., & Docking, S. I. (2016). Revisiting the continuum model of tendon pathology: what is its merit in clinical practice and research? *Br J Sports Med*, bjsports-2015-095422.

Cornelissen, V. A., & Fagard, R. H. (2005). Effects of endurance training on blood pressure, blood pressure–regulating mechanisms, and cardiovascular risk factors. *Hypertension, 46*(4), 667-675.

Corrao, G., Zambon, A., Bertù, L., Mauri, A., Paleari, V., Rossi, C., & Venegoni, M. (2006). Evidence of tendinitis provoked by fluoroquinolone treatment. *Drug safety, 29*(10), 889-896.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Daniels, D. L., & Weis, W. I. (2005). β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature structural & molecular biology, 12*(4), 364-371.

De Paoli, F., Copin, C., Vanhoutte, J., Derudas, B., Vinod, M., Zawadzki, C., . . . Staels, B. (2016). Transducin-like enhancer of split-1 is expressed and functional in human macrophages. *FEBS letters, 590*(1), 43-52.

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods, 10*(1), 5-6.

Department of Health. (2012). *Australia's Physical Activity and Sedentary Behaviour Guidelines*. Canberra.

Dias Lopes, A., Hespanhol Junior, L. C., Yeung, S. S., & Pena Costa, L. O. (2012). What are the Main Running-Related Musculoskeletal Injuries? *Sports medicine, 42*(10), 891-905.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., . . . Lecoeur, C. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. *Nature genetics, 39*(6), 724.

DNA_Genotek. (2015). Oragene DNA collection kits. Retrieved from http://www.dnagenotek.com/US/products/dnacollectionkits.html

Domaschenz, R., Vlahovich, N., Keogh, J., Compton, S., & Hughes, C. D. (2015). Exercise-Induced Tendon and Bone Injury in Recreational Runners: A Test-Retest Reliability Study. *JMIR Res Protoc, 4*(4), e117. doi:10.2196/resprot.4585

Donnelly, P. (2011). Making Sense of the Data. *Science, 331*(6020), 1024-1025. doi:10.1126/science.1204089

Dugan, S. A., & Bhat, K. P. (2005). Biomechanics and analysis of running gait. *Physical Medicine and Rehabilitation Clinics, 16*(3), 603-621.

Ekholm, O., Strandberg-Larsen, K., & Grønbæk, M. (2011). Influence of the recall period on a beverage-specific weekly drinking measure for alcohol intake. *European journal of clinical nutrition, 65*(4), 520.

El Khoury, L., Posthumus, M., Collins, M., Handley, C. J., Cook, J., & Raleigh, S. M. (2013). Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: A genetic association study. *Journal of Science and Medicine in Sport, 16*(6), 493-498.

El Khoury, L., Posthumus, M., Collins, M., van der Merwe, W., Handley, C., Cook, J. L., & Raleigh, S. M. (2015). ELN and FBN2 Gene Variants as Risk Factors for Two Sports-related Musculoskeletal Injuries. *Int J Sports Med, 36*(4), 333-337. doi:10.1055/s-0034-1390492

El Khoury, L., Ribbans, W. J., & Raleigh, S. M. (2016). MMP3 and TIMP2 gene variants as predisposing factors for Achilles tendon pathologies: Attempted replication study in a British case-control cohort. *Meta Gene, 9*, 52-55. doi:10.1016/j.mgene.2016.03.007

Elefteriou, F. (2008). Regulation of bone remodeling by the central and peripheral nervous system. *Archives of biochemistry and biophysics, 473*(2), 231-236.

Encyclopaedia Britannica. (2017). Osteocyte cell. Retrieved from https://www.britannica.com/science/osteocyte

Engelen, L., Gale, J., Chau, J. Y., Hardy, L. L., Mackey, M., Johnson, N., . . . Bauman, A. (2016). Who is at risk of chronic disease? Associations between risk profiles of physical activity, sitting and cardio-metabolic disease in Australian adults. *Australian and New Zealand Journal of Public Health*.

Erhardt, L. (2009). Cigarette smoking: an undertreated risk factor for cardiovascular disease. *Atherosclerosis, 205*(1), 23-32.

Esculier, J.-F., Bouyer, L. J., Dubois, B., Fremont, P., Moore, L., McFadyen, B., & Roy, J.-S. (2017). Is combining gait retraining or an exercise programme with education better than education alone in treating runners with patellofemoral pain? A randomised clinical trial. *Br J Sports Med*, bjsports-2016-096988.

Esen, İ., Demirel, F., Güven, A., Değerliyurt, A., & Köse, G. (2011). Assessment of bone density in children with cerebral palsy by areal bone mineral density measurement. *measurement, 53*, 638-644.

Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y.-H., Duncan, E. L., Ntzani, E. E., . . . Kemp, J. P. (2012). Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature genetics, 44*(5), 491-501.

Evangelou, E., & Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics, 14*(6), 379.

Farris, D. J., Buckeridge, E., Trewartha, G., & McGuigan, M. P. (2012). The effects of orthotic heel lifts on Achilles tendon force and strain during running. *J Appl Biomech, 28*(5), 511-519.

Fechtenbaum, M., Nam, J. L., & Emery, P. (2014). Biologics in rheumatoid arthritis: where are we going? *Br J Hosp Med (Lond), 75*, 448-449.

Ferrari, S., Rizzoli, R., Slosman, D., & Bonjour, J. P. (1998). Do dietary calcium and age explain the controversy surrounding the relationship between bone mineral density and vitamin D receptor gene polymorphisms? *Journal of Bone and Mineral Research, 13*(3), 363-370.

Fishman, D., Faulds, G., Jeffery, R., Mohamed-Ali, V., Yudkin, J. S., Humphries, S., & Woo, P. (1998). The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *Journal of Clinical Investigation, 102*(7), 1369.

Flinn, S. D. (2002). Changes in stress fracture distribution and current treatment. *Curr Sports Med Rep, 1*(5), 272-277.

Fogelholm, M. (2010). Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. A systematic review. *Obesity reviews, 11*(3), 202-221.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., . . . Rayner, N. W. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science, 316*(5826), 889-894.

Fredericson, M., Cookingham, C. L., Chaudhari, A. M., Dowdell, B. C., Oestreicher, N., & Sahrmann, S. A. (2000). Hip abductor weakness in distance runners with iliotibial band syndrome. *Clinical Journal of Sport Medicine, 10*(3), 169-175.

Gabbe, B. J., Finch, C. F., Bennell, K. L., & Wajswelner, H. (2003). How valid is a self reported 12 month sports injury history? *British Journal of Sports Medicine, 37*(6), 545-547.

Gaida, J. E., Alfredson, H., Kiss, Z. S., Bass, S. L., & Cook, J. L. (2010). Asymptomatic Achilles tendon pathology is associated with a central fat distribution in men and a peripheral fat distribution in women: a cross sectional study of 298 individuals. *BMC Musculoskeletal Disorders, 11*, 41-41. doi:10.1186/1471-2474-11-41

Gaida, J. E., Alfredson, L., Kiss, Z. S., Wilson, A. M., Alfredson, H., & Cook, J. L. (2009). Dyslipidemia in achilles tendinopathy is characteristic of insulin resistance. *Medicine and science in sports and exercise, 41*(6), 1194-1197.

Gaida, J. E., Bagge, J., Purdam, C. R., Cook, J. L., Alfredson, H., & Forsgren, S. (2012). Evidence of the TNF-α System in the Human Achilles Tendon: Expression of TNF-α and TNF Receptor at both Protein and mRNA Levels in the Tenocytes. *Cells Tissues Organs, 196*(4), 339-352.

Gibbon, A., Hobbs, H., van der Merwe, W., Raleigh, S. M., Cook, J., Handley, C. J., . . . September, A. V. (2016). The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups. *Journal of Sports Sciences*, 1-8.

Girjes, A. A., Hobson, K., Chen, P., & Lavin, M. F. (1995). Cloning and characterization of cDNA encoding a human arginyl-tRNA synthetase. *Gene, 164*(2), 347-350.

Grier, T., Canham-Chervak, M., Bushman, T., Anderson, M., North, W., & Jones, B. H. (2016). Minimalist running shoes and injury risk among United States army soldiers. *The American journal of sports medicine, 44*(6), 1439-1446.

Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the practice of exercise and sport-science research. *International journal of sports physiology and performance, 13*(2), 127-134.

Haplotype Reference Consortium. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics, 48*(10), 1279-1283.

Hart, L. E. (1994). Exercise and soft tissue injury. *Baillieres Clin Rheumatol, 8*(1), 137-148.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116): Sinauer associates Sunderland.

Håvik, B., Le Hellard, S., Rietschel, M., Lybæk, H., Djurovic, S., Mattheisen, M., . . . Maier, W. (2011). The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biological psychiatry, 70*(1), 35-42.

Hay, M., Patricios, J., Collins, R., Branfield, A., Cook, J. L., Handley, C. J., . . . Collins, M. (2013). Association of type XI collagen genes with chronic Achilles tendinopathy in independent populations from South Africa and Australia. *Br J Sports Med, 47*(9), 569-574. doi:10.1136/bjsports-2013-092379

Hespanhol Junior, L. C., Costa, L. O. P., & Lopes, A. D. (2013). Previous injuries and some training characteristics predict running-related injuries in recreational runners: a prospective cohort study. *Journal of Physiotherapy, 59*(4), 263-269.

Hespanhol Junior, L. C., Pillay, J. D., van Mechelen, W., & Verhagen, E. (2015). Meta-analyses of the effects of habitual running on indices of health in physically inactive adults. *Sports medicine, 45*(10), 1455-1468.

Hildebrand, J. K., Maria. Majewski, Ian. Liu, Zikou. Cox, Allison. Miyake, Sanae. Hall, Cathrine. Petrie, Emma. Silk, Michael. Tanzer, Maria. Young, Samuel. Garnish, Sarah. Corbin, Jason. Stutz, Michael. Gangatirkar, Pradnya. Josefsson, Emma. Rigbye, Kristin. Anderton, Holly. Rickard, James. Silke, John. . (2019). Missense mutations in the MLKL brace region lead to lethal neonatal inflammation in mice and are present in high frequency in humans. [Under review]. *Nature Cell Biology*. doi:10.1101/628370

Hirabayashi, R., Hozumi, S., Higashijima, S.-i., & Kikuchi, Y. (2013). Ddx46 is required for multi-lineage differentiation of hematopoietic stem cells in zebrafish. *Stem cells and development, 22*(18), 2532-2542.

Hirschmüller, A., Frey, V., Konstantinidis, L., Baur, H., Dickhuth, H.-H., Südkamp, N. P., & Helwig, P. (2012). Prognostic value of Achilles tendon Doppler sonography in asymptomatic runners. *Medicine and science in sports and exercise, 44*(2), 199-205. doi:10.1249/MSS.0b013e31822b7318

Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & informatics, 10*(2), 117-122.

Houlihan, C. M., & Stevenson, R. D. (2009). Bone density in cerebral palsy. *Physical medicine and rehabilitation clinics of North America, 20*(3), 493-508.

Hozumi, S., Hirabayashi, R., Yoshizawa, A., Ogata, M., Ishitani, T., Tsutsumi, M., . . . Kikuchi, Y. (2012). DEAD-box protein Ddx46 is required for the development of the digestive organs and brain in zebrafish. *PloS one, 7*(3), e33675.

Hreljac, A. (2005). Etiology, prevention, and early intervention of overuse injuries in runners: a biomechanical perspective. *Physical Medicine and Rehabilitation Clinics, 16*(3), 651-667.

Hreljac, A., Marshall, R. N., & Hume, P. A. (2000). Evaluation of lower extremity overuse injury potential in runners. *Medicine & Science in Sports & Exercise, 32*(9), 1635-1641.

Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: roles in neuronal development and function. *Annual review of neuroscience, 24*(1), 677-736.

Ichikawa, S., Fukuhara, N., Katsushima, H., Takahashi, T., Yamamoto, J., Yokoyama, H., . . . Ishizawa, K. (2014). Association between BACH2 expression and clinical prognosis in diffuse large B-cell lymphoma. *Cancer science, 105*(4), 437-444.

Ichtchenko, K., Hata, Y., Nguyen, T., Ullrich, B., Missler, M., Moomaw, C., & Südhof, T. C. (1995). Neuroligin 1: a splice site-specific ligand for β-neurexins. *Cell, 81*(3), 435-443.

Illumina. (2015). Infinium CoreExome-24 BeadChip Retrieved from http://www.illumina.com/products/humancore_exome_beadchip_kits.html

Inhofe, P. D., Grana, W. A., Egle, D., Min, K.-W., & Tomasek, J. (1995). The Effects of Anabolic Steroids on Rat Tendon An Ultrastructural, Biomechanical, and Biochemical Analysis. *The American journal of sports medicine, 23*(2), 227-232.

Ismail, I., Keating, S., Baker, M., & Johnson, N. (2012). A systematic review and meta-analysis of the effect of aerobic vs. resistance exercise training on visceral fat. *Obesity reviews, 13*(1), 68-91.

Järvinen, T. A., Kannus, P., Maffulli, N., & Khan, K. M. (2005). Achilles tendon disorders: etiology and epidemiology. *Foot and ankle clinics, 10*(2), 255-266.

Jiang, F., Zhang, D., Li, G., & Wang, X. (2017). Knockdown of DDX46 Inhibits the Invasion and Tumorigenesis in Osteosarcoma Cells. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics, 25*(3), 417-425.

Kaczynski, A. T., Manske, S. R., Mannell, R. C., & Grewal, K. (2008). Smoking and physical activity: a systematic review. *American journal of health behavior, 32*(1), 93-110.

Kader, D., Saxena, A., Movin, T., & Maffulli, N. (2002). Achilles tendinopathy: some aspects of basic science and clinical management. *British Journal of Sports Medicine, 36*(4), 239.

Kanis, J. A., Johansson, H., Oden, A., Johnell, O., De Laet, C., Melton, L. J., . . . Pols, H. A. (2004). A meta-analysis of prior corticosteroid use and fracture risk. *Journal of Bone and Mineral Research, 19*(6), 893-899.

Kanis, J. A., Johnell, O., Odén, A., Johansson, H., De Laet, C., Eisman, J. A., . . . Mellstrom, D. (2005). Smoking and fracture risk: a meta-analysis. *Osteoporosis international, 16*(2), 155-162.

Karlsson, M., & Rosengren, B. (2012). Training and bone–from health to injury. *Scandinavian Journal of Medicine & Science in Sports, 22*(4), e15-e23.

Katoh, M., & Katoh, M. (2004). Identification and characterization of human PDZRN4L gene and mouse Pdzrn4l gene in silico. *International journal of molecular medicine, 13*(6), 923-927.

Kehoe, A., & Montgomery, H. (2006). Genetic variation and the skeletal response to exercise: a systematic review: review article. *International SportMed Journal: Genetics and Sport, 7*(3), p. 187-200.

Khaliq, Y., & Zhanel, G. G. (2003). Fluoroquinolone-associated tendinopathy: a critical review of the literature. *Clinical infectious diseases, 36*(11), 1404-1410.

Kim, S. K., Roos, T. R., Roos, A. K., Kleimeyer, J. P., Ahmed, M. A., Goodlin, G. T., . . . Dragoo, J. L. (2017). Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PloS one, 12*(3), e0170422.

Kiuru, M. J., Pihlajamäki, H., & Ahovuo, J. (2004). Bone stress injuries. *Acta Radiologica, 45*(3), 000-000.

Kjaer, M., Bayer, M. L., Eliasson, P., & Heinemeier, K. M. (2013). What is the impact of inflammation on the critical interplay between mechanical signaling and biochemical changes in tendon matrix? *Journal of Applied Physiology, 115*(6), 879-883.

Klein, C., Lohmann, K., & Ziegler, A. (2012). The promise and limitations of genome-wide association studies. *Jama, 308*(18), 1867-1868.

Knobloch, K., Yoon, U., & Vogt, P. M. (2008). Acute and overuse injuries correlated to hours of training in master running athletes. *Foot Ankle Int, 29*(7), 671-676.

Kolt, G. S., & Kirkby, R. J. (1999). Epidemiology of injury in elite and subelite female gymnasts: a comparison of retrospective and prospective findings. *British Journal of Sports Medicine, 33*(5), 312-318.

Korpelainen, R., Orava, S., Karpakka, J., Siira, P., & Hulkko, A. (2001a). Risk factors for recurrent stress fractures in athletes. *The American journal of sports medicine, 29*(3), 304-310.

Korpelainen, R., Orava, S., Karpakka, J., Siira, P., & Hulkko, A. (2001b). Risk Factors for Recurrent Stress Fractures in Athletes No author or related institution has received any financial benefit from research in this study. See "Acknowledgment" for funding information. *The American journal of sports medicine, 29*(3), 304-310.

Korsten-Reck, U. (2011). The female athlete triad: FIMS position statement 2011. *International SportMed Journal, 12*(4), 156-159.

Kraemer, R., Wuerfel, W., Lorenzen, J., Busche, M., Vogt, P., & Knobloch, K. (2012). Analysis of hereditary and medical risk factors in Achilles tendinopathy and Achilles tendon ruptures: a matched pair analysis. *Archives of Orthopaedic & Trauma Surgery, 132*(6), 847-853.

Krishnan, V., Bryant, H. U., & MacDougald, O. A. (2006). Regulation of bone mass by Wnt signaling. *The Journal of clinical investigation, 116*(5), 1202-1209.

Kruijshaar, M. E., Barendregt, J., Vos, T., De Graaf, R., Spijker, J., & Andrews, G. (2005). Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. *European journal of epidemiology, 20*(1), 103-111.

Kulms, D., & Schwarz, T. (2006). NF-κB and Cytokines. *Vitamins & Hormones, 74*, 283-300.

Kuo, K.-T., Guan, B., Feng, Y., Mao, T.-L., Chen, X., Jinawath, N., . . . Wang, T.-L. (2009). Analysis of DNA copy number alterations in ovarian serous tumors identifies new molecular genetic changes in low-grade and high-grade carcinomas. *Cancer research, 69*(9), 4036-4042.

Kuwahara, M., Ise, W., Ochi, M., Suzuki, J., Kometani, K., Maruyama, S., . . . Takemori, A. (2016). Bach2–Batf interactions control Th2-type immune response by regulating the IL-4 amplification loop. *Nature communications, 7*, 12596.

Landvik, N. E., Hart, K., Skaug, V., Stangeland, L. B., Haugen, A., & Zienolddiny, S. (2009). A specific interleukin-1B haplotype correlates with high levels of IL1B mRNA in the lung and increased risk of non-small cell lung cancer. *Carcinogenesis, 30*(7), 1186-1192.

Langdahl, B. L., Løkke, E., Carstens, M., Stenkjær, L. L., & Eriksen, E. F. (2000). Osteoporotic Fractures Are Associated with an 86-Base Pair Repeat Polymorphism in the Interleukin-1-Receptor Antagonist Gene But Not with Polymorphisms in the Interleukin-1β Gene. *Journal of Bone and Mineral Research, 15*(3), 402-414.

Lappe, J., Cullen, D., Haynatzki, G., Recker, R., Ahlf, R., & Thompson, K. (2008). Calcium and vitamin D supplementation decreases incidence of stress fractures in female navy recruits. *Journal of Bone and Mineral Research, 23*(5), 741-749.

Lappe, J., Stegman, M., & Recker, R. (2001). The impact of lifestyle factors on stress fractures in female Army recruits. *Osteoporosis international, 12*(1), 35-42.

Laseter, J. T., & Russell, J. A. (1991). Anabolic steroid-induced tendon pathology: a review of the literature. *Medicine and science in sports and exercise, 23*(1), 1-3.

Lauderdale, M. E., Yli-Piipari, S., Irwin, C. C., & Layne, T. E. (2015). Gender differences regarding motivation for physical activity among college students: A self-determination approach. *The Physical Educator, 72*(1), 153-172.

Law, M. H., Choi, E. M., Law, S. H., Chan, S. S., Wong, S. M., Ching, E. C., . . . Lau, F. O. (2018). Effects of footwear midsole thickness on running biomechanics. *Journal of Sports Sciences*, 1-7.

Lee, D. C., Pate, R. R., Lavie, C. J., Sui, X., Church, T. S., & Blair, S. N. (2014). Leisure-time running reduces all-cause and cardiovascular mortality risk. *Journal of the American College of Cardiology, 64*(5), 472-481.

Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., & Group, L. P. A. S. W. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet, 380*(9838), 219-229.

Leppilahti, J., Puranen, J., & Orava, S. (1995). *ABO blood group and Achilles tendon rupture.* Paper presented at the Annales chirurgiae et gynaecologiae.

Lewis, C. M., Ng, M. Y., Butler, A. W., Cohen-Woods, S., Uher, R., Pirlo, K., . . . Rivera, M. (2010). Genome-wide association study of major recurrent depression in the UK population. *American Journal of Psychiatry, 167*(8), 949-957.

Li, W., Soave, D., Miller, M. R., Keenan, K., Lin, F., Gong, J., . . . Rommens, J. (2014). Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. *Human genetics, 133*(2), 151-161.

Li, Z., Yang, G., Khan, M., Stone, D., Woo, S. L., & Wang, J. H. (2004). Inflammatory response of human tendon fibroblasts to cyclic mechanical stretching. *The American journal of sports medicine, 32*(2), 435-440.

Linton, L., & Valentin, S. (2018). Running with injury: a study of UK novice and recreational runners and factors associated with running related injury. *Journal of Science and Medicine in Sport*.

Loef, M., & Walach, H. (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Preventive medicine, 55*(3), 163-170.

Longo, U. G., Rittweger, J., Garau, G., Radonic, B., Gutwasser, C., Gilliver, S. F., . . . Maffulli, N. (2009a). No influence of age, gender, weight, height, and impact profile in achilles tendinopathy in masters track and field athletes. *The American journal of sports medicine, 37*(7), 1400-1405.

Longo, U. G., Ronga, M., & Maffulli, N. (2009b). Achilles Tendinopathy. *Sports Medicine & Arthroscopy Review, 17*(2), 112-126.

Loos, R. J., & Yeo, G. S. (2014). The bigger picture of FTO—the first GWAS-identified obesity gene. *Nature Reviews Endocrinology, 10*(1), 51.

Lorimer, A., & Hume, P. (2014). Achilles Tendon Injury Risk Factors Associated with Running. *Sports medicine, 44*(10), 1459-1472.

Loughner, C. L., Bruford, E. A., McAndrews, M. S., Delp, E. E., Swamynathan, S., & Swamynathan, S. K. (2016). Organization, evolution and functions of the human and mouse Ly6/uPAR family genes. *Human genomics, 10*(1), 10.

Louw, A., Van Biljon, A., & Mugandani, S. (2012). Exercise motivation and barriers among men and women of different age groups: sport psychology. *African Journal for Physical Health Education, Recreation and Dance, 18*(Issue-4_1), 759-768.

MacAuley, D. (1994). A history of physical activity, health and medicine. *Journal of the Royal Society of Medicine, 87*(1), 32.

Macera, C. A., Pate, R. R., Powell, K. E., Jackson, K. L., Kendrick, J. S., & Craven, T. E. (1989). Predicting lower-extremity injuries among habitual runners. *Archives of internal medicine, 149*(11), 2565-2568.

Maffulli, N. (1998). Overuse tendon conditions: time to change a confusing terminology. *Arthroscopy: The Journal of Arthroscopic & Related Surgery, 14*(8), 840-843.

Maffulli, N., Reaper, J. A., Waterston, S. W., & Ahya, R. (2000). ABO blood groups and Achilles tendon rupture in the Grampian region of Scotland. *Clinical Journal of Sport Medicine, 10*(4), 269-271.

Maffulli, N., Wong, J., & Almekinders, L. C. (2003). Types and epidemiology of tendinopathy. *Clinics in sports medicine, 22*(4), 675-692.

Magnan, B., Bondi, M., Pierantoni, S., & Samaila, E. (2014). The pathogenesis of Achilles tendinopathy: a systematic review. *Foot And Ankle Surgery: Official Journal Of The European Society Of Foot And Ankle Surgeons, 20*(3), 154-159. doi:10.1016/j.fas.2014.02.010

Mahler, F., & Fritschy, D. (1992). Partial and complete ruptures of the Achilles tendon and local corticosteroid injections. *British Journal of Sports Medicine, 26*(1), 7-14.

Manolagas, S. (1995). Role of cytokines in bone resorption. *Bone, 17*(2), S63-S67.

Manolagas, S. C. (2000). Birth and death of bone cells: basic regulatory mechanisms and implications for the pathogenesis and treatment of osteoporosis. *Endocrine reviews, 21*(2), 115-137.

Manolio, T. A., Brooks, L. D., & Collins, F. S. (2008). A HapMap harvest of insights into the genetics of common disease. *The Journal of clinical investigation, 118*(5), 1590-1605.

Mansfield, J. C., Holden, H., Tarlow, J. K., Di Giovine, F. S., McDowell, T. L., Wilson, A. G., . . . Duff, G. W. (1994). Novel genetic association between ulcerative colitis and the anti-inflammatory cytokine interleukin-1 receptor antagonist. *Gasrtoenterology-Baltimore then Philadelphia, 106*, 637-637.

Manzanero, S., Kozlovskaia, M., Vlahovich, N., & Hughes, D. C. (2018). Recruitment and Participation of Recreational Runners in a Large Epidemiological and Genetic Research Study: Retrospective Data Analysis. *JMIR research protocols, 7*(5).

Mattila, V. M., Niva, M., Kiuru, M., & Pihlajamäki, H. (2007). Risk factors for bone stress injuries: a follow-up study of 102,515 person-years. *Medicine and science in sports and exercise, 39*(7), 1061-1066.

McCrory, J. L., Martin, D. F., Lowery, R. B., Cannon, D. W., Curl, W. W., Read Jr, H. M., . . . Messier, S. P. (1999). Etiologic factors associated with Achilles tendinitis in runners. *Medicine and science in sports and exercise, 31*(10), 1374-1381.

McIlwain, C., Townsend, D., & Tew, K. (2006). Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene, 25*(11), 1639.

McKean, K. A., Manson, N. A., & Stanish, W. D. (2006). Musculoskeletal injury in the masters runners. *Clinical Journal of Sport Medicine, 16*(2), 149-154.

McKenzie, D., Clement, D., & Taunton, J. (1985). Running shoes, orthotics, and injuries. *Sports medicine, 2*(5), 334-347.

McQuillan, G. M., & Porter, K. S. (2011). Consent for future genetic research: the NHANES experience in 2007-2008. *Irb, 33*(1), 9.

Melzer, K., Kayser, B., & Pichard, C. (2004). Physical activity: the health benefits outweigh the risks. *Current Opinion in Clinical Nutrition & Metabolic Care, 7*(6), 641-647.

Memet, I., Doebele, C., Sloan, K. E., & Bohnsack, M. T. (2017). The G-patch protein NF-κB-repressing factor mediates the recruitment of the exonuclease XRN2 and activation of the RNA helicase DHX15 in human ribosome biogenesis. *Nucleic acids research, 45*(9), 5359-5374.

Messier, S. P., Martin, D. F., Mihalko, S. L., Ip, E., DeVita, P., Cannon, D. W., . . . Fellin, R. E. (2018). A 2-Year Prospective Cohort Study of Overuse Running Injuries: The Runners and Injury Longitudinal Study (TRAILS). *The American journal of sports medicine*, 0363546518773755.

Mezuk, B., Eaton, W. W., & Zandi, P. (2008). Participant characteristics that influence consent for genetic research in a population-based survey: the Baltimore epidemiologic catchment area follow-up. *Public Health Genomics, 11*(3), 171-178.

Michaud, L. B., & Goodin, S. (2006). Cancer-treatment-induced bone loss, part 1. *American journal of health-system pharmacy, 63*(5), 419-430.

Moayyeri, A., Hsu, Y.-H., Karasik, D., Estrada, K., Xiao, S.-M., Nielson, C., . . . Zheng, H.-F. (2014). Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. *Human molecular genetics, 23*(11), 3054-3068.

Mokone, G. G., Gajjar, M., September, A. V., Schwellnus, M. P., Greenberg, J., Noakes, T. D., & Collins, M. (2005). The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. *American Journal of Sports Medicine, 33*(7), 1016-1021.

Mokone, G. G., Schwellnus, M. P., Noakes, T. D., & Collins, M. (2006). The COL5A1 gene and Achilles tendon pathology. *Scandinavian Journal of Medicine & Science in Sports, 16*(1), 19-26.

Morales, J. C., Richard, P., Patidar, P. L., Motea, E. A., Dang, T. T., Manley, J. L., & Boothman, D. A. (2016). XRN2 links transcription termination to DNA damage and replication stress. *PLoS genetics, 12*(7), e1006107.

Morales, M. G., Cabrera, D., Céspedes, C., Vio, C. P., Vazquez, Y., Brandan, E., & Cabello-Verrugio, C. (2013). Inhibition of the angiotensin-converting enzyme decreases skeletal muscle fibrosis in dystrophic mice by a diminution in the expression and activity of connective tissue growth factor (CTGF/CCN-2). *Cell and tissue research, 353*(1), 173-187.

Morales, M. G., Vazquez, Y., Acuña, M. J., Rivera, J. C., Simon, F., Salas, J. D., . . . Cabello-Verrugio, C. (2012). Angiotensin II-induced pro-fibrotic effects require p38MAPK activity and transforming growth factor beta 1 expression in skeletal muscle cells. *The international journal of biochemistry & cell biology, 44*(11), 1993-2002.

Moriarity, B. S., Otto, G. M., Rahrmann, E. P., Rathe, S. K., Wolf, N. K., Weg, M. T., . . . Molyneux, S. D. (2015). A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nature genetics, 47*(6), 615.

Mountjoy, M., Sundgot-Borgen, J., Burke, L., Carter, S., Constantini, N., Lebrun, C., . . . Budgett, R. (2015). Authors' 2015 additions to the IOC consensus statement: Relative Energy Deficiency in Sport (RED-S). *British Journal of Sports Medicine, 49*(7), 417-420.

Myers, J. (2003). Exercise and cardiovascular health. *Circulation, 107*(1), e2-e5.

Nakanishi, M., Nomura, J., Ji, X., Tamada, K., Arai, T., Takahashi, E., . . . Takumi, T. (2017). Functional significance of rare neuroligin 1 variants found in autism. *PLoS genetics, 13*(8), e1006940.

Namipashaki, A., Razaghi-Moghadam, Z., & Ansari-Pour, N. (2015). The essentiality of reporting Hardy-Weinberg equilibrium calculations in population-based genetic association studies. *Cell Journal (Yakhteh), 17*(2), 187.

Nampei, A., Hashimoto, J., Koyanagi, J., Ono, T., Hashimoto, H., Tsumaki, N., . . . Ochi, T. (2008). Characteristics of fracture and related factors in patients with rheumatoid arthritis. *Modern Rheumatology, 18*(2), 170-176.

National Heart, L. a. B. I. (2017). Full genomic results download - Kim et al. Achilles summary. Retrieved from https://grasp.nhlbi.nih.gov/FullResults.aspx

Nell, E. M., Van Der Merwe, L., Cook, J. L., Handley, C. J., Collins, M., & September, A. (2012). The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. *Journal of Orthopaedic Research, 30*(11), 1719-1724.

Newman, P., Witchalls, J., Waddington, G., & Adams, R. (2013). Risk factors associated with medial tibial stress syndrome in runners: a systematic review and meta-analysis. *Open access journal of sports medicine, 4*, 229.

Nielsen, R. O., Ronnow, L., Rasmussen, S., & Lind, M. (2014). A prospective study on time to recovery in 254 injured novice runners. *PloS one, 9*(6), e99877. doi:10.1371/journal.pone.0099877

Niemuth, P. E., Johnson, R. J., Myers, M. J., & Thieman, T. J. (2005). Hip muscle weakness and overuse injuries in recreational runners. *Clinical Journal of Sport Medicine, 15*(1), 14-21.

Nieves, J. W., Melsop, K., Curtis, M., Kelsey, J. L., Bachrach, L. K., Greendale, G., . . . Sainani, K. L. (2010). Nutritional factors that influence change in bone density and stress fracture risk among young female cross-country runners. *PM&R, 2*(8), 740-750.

Nigg, B. M., Nurse, M. A., & Stefanyshyn, D. J. (1999). Shoe inserts and orthotics for sport and physical activities. *Medicine and science in sports and exercise, 31*, S421-S428.

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current directions in psychological science, 10*(5), 173-176.

O'keefe, J. H., & Lavie, C. J. (2013). Run for your life… at a comfortable speed and not too far. In: BMJ Publishing Group Ltd and British Cardiovascular Society.

Ohkawara, K., Tanaka, S., Miyachi, M., Ishikawa-Takata, K., & Tabata, I. (2007). A dose–response relation between aerobic exercise and visceral fat reduction: systematic review of clinical trials. *International journal of obesity, 31*(12), 1786-1797.

Oja, P., Kelly, P., Pedisic, Z., Titze, S., Bauman, A., Foster, C., . . . Stamatakis, E. (2016). Associations of specific types of sports and exercise with all-cause and cardiovascular-disease mortality: a cohort study of 80 306 British adults. *British Journal of Sports Medicine*, bjsports-2016-096822.

Oja, P., Titze, S., Kokko, S., Kujala, U. M., Heinonen, A., Kelly, P., . . . Foster, C. (2015). Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis. *British Journal of Sports Medicine*, bjsports-2014-093885.

Olofsson, P. S., Sheikine, Y., Jatta, K., Ghaderi, M., Samnegård, A., Eriksson, P., & Sirsjö, A. (2009). A Functional Interleukin-1 Receptor Antagonist Polymorphism Influences Atherosclerosis Development The Interleukin-1. BETA.: Interleukin-1 Receptor Antagonist Balance in Atherosclerosis. *Circulation Journal, 73*(8), 1531-1536.

Ortega, S.-P. R., & Aguilar-Blanco, E. (2006). [Running and its influence on smoking habits]. *Atencion primaria/Sociedad Espanola de Medicina de Familia y Comunitaria, 37*(9), 478-481.

Owens, B. D., Wolf, J. M., Seelig, A. D., Jacobson, I. G., Boyko, E. J., Smith, B., . . . Smith, T. C. (2013). Risk factors for lower extremity tendinopathies in military personnel. *Orthopaedic journal of sports medicine, 1*(1), 1-8.

Pack, A. M. (2003). The association between antiepileptic drugs and bone disease. *Epilepsy Currents, 3*(3), 91-95.

Padhiar, N., Acharya, N., Chan, R., Davinii, K., Crisp, T., King, J., . . . Webborn, N. (2010). Achilles tendinopathy Part 1 - PAthophysiology and clinical features. *SportEX Medicine*(45), 23-30.

Panagopoulos, V. N., Trull, T. J., Glowinski, A. L., Lynskey, M. T., Heath, A. C., Agrawal, A., . . . Madden, P. A. (2013). Examining the association of NRXN3 SNPs with borderline personality disorder phenotypes in heroin dependent cases and socio-economically disadvantaged controls. *Drug and alcohol dependence, 128*(3), 187-193.

Pandya, C. D., Kutiyanawalla, A., & Pillai, A. (2013). BDNF–TrkB signaling and neuroprotection in schizophrenia. *Asian journal of psychiatry, 6*(1), 22-28.

Paramonov, A., Kulbatskii, D., Loktyushov, E., Tsarev, A., Dolgikh, D., Shenkarev, Z., . . . Lyukmanova, E. (2017). Recombinant production and structural studies of the human Lypd6 and Lypd6b proteins. *Russian Journal of Bioorganic Chemistry, 43*(6), 644-652.

Pećina-Šlaus, N., Kafka, A., & Lechpammer, M. (2016). Molecular genetics of intracranial meningiomas with emphasis on canonical Wnt signalling. *Cancers, 8*(7), 67.

Peters, J. A., Zwerver, J., Diercks, R. L., Elferink-Gemser, M. T., & van den Akker-Scheek, I. (2015). Preventive interventions for tendinopathy: A systematic review. *Journal of Science and Medicine in Sport*.

Petty, S. J., O'Brien, T., & Wark, J. (2007). Anti-epileptic medication and bone health. *Osteoporosis international, 18*(2), 129-142.

Pfeiffer, R. M., & Gail, M. H. (2003). Sample size calculations for population-and family-based case-control association studies on marker genotypes. *Genetic epidemiology, 25*(2), 136-148.

Physical Activity Guidelines Advisory Committee. (2008). Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services, 2008*, A1-H14.

Piters, E., Boudin, E., & Van Hul, W. (2008). Wnt signaling: a win for bone. *Archives of biochemistry and biophysics, 473*(2), 112-116.

Pitsiladis, Y. P., Tanaka, M., Eynon, N., Bouchard, C., North, K. N., Williams, A. G., . . . Fuku, N. (2016). Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiological genomics, 48*(3), 183-190.

Pollock, N. (2011). Stress fractures in sport. *SportEX Medicine*(50).

Population Australia. (2018). Australia population 2018. Retrieved from http://www.population.net.au/

Porter, A. G., & Jänicke, R. U. (1999). Emerging roles of caspase-3 in apoptosis. *Cell death and differentiation, 6*(2), 99-104.

Porter, S., Eccleston, P., & Vilshanskaya, O. (2002). Moving patients towards a more active lifestyle: the GP Physical Activity Project in South Eastern Sydney Area Health Service. *Health Promotion Journal of Australia, 13*(3), 178-183.

Posthumus, M., Collins, M., Cook, J., Handley, C. J., Ribbans, W. J., Smith, R. K., . . . Raleigh, S. M. (2010). Components of the transforming growth factor-β family and the pathogenesis of human Achilles tendon pathology—a genetic association study. *Rheumatology, 49*(11), 2090-2097.

Power-Calculator, C. (2015). Retrieved from http://csg.sph.umich.edu/abecasis/CaTS/interface.html

Protzman, R., & Griffis, C. (1977). Stress fractures in men and women undergoing military training. *The Journal of Bone & Joint Surgery, 59*(6), 825-825.

Proude, E. M., Britt, H., Valenti, L., & Conigrave, K. M. (2006). The relationship between self-reported alcohol intake and the morbidities managed by GPs in Australia. *BMC family practice, 7*(1), 17.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., . . . Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26*(18), 2336-2337.

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics, 19*(1), 149-150.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

R Development Core Team. (2008). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Raisz, L. G. (2005). Pathogenesis of osteoporosis: concepts, conflicts, and prospects. *Journal of Clinical Investigation, 115*(12), 3318.

Raleigh, S. M. (2009). Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. *British Journal of Sports Medicine, 43*(7), 514-520.

Ray, M., Weickert, C. S., & Webster, M. (2014). Decreased BDNF and TrkB mRNA expression in multiple cortical areas of patients with schizophrenia and mood disorders. *Translational psychiatry, 4*(5), e389.

Rehman, Q., & Lane, N. E. (2003). Effect of glucocorticoids on bone density. *Medical and pediatric oncology, 41*(3), 212-216.

Reiner, M., Niermann, C., Jekauc, D., & Woll, A. (2013). Long-term health benefits of physical activity–a systematic review of longitudinal studies. *BMC public health, 13*(1), 813.

Reissner, C., Klose, M., Fairless, R., & Missler, M. (2008). Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. *Proceedings of the National Academy of Sciences, 105*(39), 15124-15129.

Richards, C. E., Magin, P. J., & Callister, R. (2009). Is your prescription of distance running shoes evidence-based? *British Journal of Sports Medicine, 43*(3), 159-162.

Richardson, C. R., Kriska, A. M., Lantz, P. M., & Hayward, R. A. (2004). Physical activity and mortality across cardiovascular disease risk groups. *Medicine and science in sports and exercise, 36*(11), 1923-1929.

Rickaby, R., El Khoury, L., Ribbans, W. J., & Raleigh, S. M. (2015). Variation within three apoptosis associated genes as potential risk factors for Achilles tendinopathy in a British based case-control cohort. *Gene, 571*(2), 167-171. doi:10.1016/j.gene.2015.06.010

Rickert, M., Wang, H., Wieloch, P., Lorenz, H., Steck, E., Sabo, D., & Richter, W. (2005). Adenovirus-mediated gene transfer of growth and differentiation factor-5 into tenocytes and the healing rat Achilles tendon. *Connective Tissue Research, 46*(4-5), 175-183.

Rixe, J. A., Gallo, R. A., & Silvis, M. L. (2012). The barefoot debate: can minimalist shoes reduce running-related injuries? *Current sports medicine reports, 11*(3), 160-165.

Robertson, J., & Macdonald, K. (2010). Prevalence of bone loss in a population with cystic fibrosis. *British Journal of Nursing, 19*(10), 636-639.

Robertson, R., Jepson, R., Shepherd, A., & McInnes, R. (2011). Recommendations by Queensland GPs to be more physically active: which patients were recommended which activities and what action they took. *Australian and New Zealand Journal of Public Health, 35*(6), 537-542.

Roper, J. L., Harding, E. M., Doerfler, D., Dexter, J. G., Kravitz, L., Dufek, J. S., & Mermier, C. M. (2016). The effects of gait retraining in runners with patellofemoral pain: A randomized trial. *Clinical biomechanics, 35*, 14-22.

Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., . . . Jacobsson, B. (2007). A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol Suppl, 109*(suppl 109), 8-14.

Rother, S., Bartels, M., Schweda, A. T., Resch, K., Pallua, N., & Nourbakhsh, M. (2016). NF-κB–repressing factor phosphorylation regulates transcription elongation via its interactions with 5'→ 3' exoribonuclease 2 and negative elongation factor. *The FASEB Journal, 30*(1), 174-185.

Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., . . . Hood, L. (2002). Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics, 79*(4), 587-597.

Rubinsztein, D. C., & Easton, D. F. (1999). Apolipoprotein E genetic variation and Alzheimer's disease. *Dementia and geriatric cognitive disorders, 10*(3), 199-209.

Ruiz-Larrañaga, O., Uribarri, M., Alcaro, M. C., Escorza-Treviño, S., Del Amo, J., Iriondo, M., . . . Estonba, A. (2016). Genetic variants associated with rheumatoid arthritis patients and serotypes in European populations. *Clin. Exp. Rheumatol., 34*, 236-241.

Ruohola, J. P., Laaksi, I., Ylikomi, T., Haataja, R., Mattila, V. M., Sahi, T., . . . Pihlajamäki, H. (2006). Association between serum 25 (OH) D concentrations and bone stress fractures in Finnish young men. *Journal of Bone and Mineral Research, 21*(9), 1483-1488.

Ryan, M., Elashi, M., Newsham-West, R., & Taunton, J. (2014). Examining injury risk and pain perception in runners using minimalist footwear. *Br J Sports Med, 48*(16), 1257-1262.

Sanger Institute. (2017). Sanger Imputation Service. Retrieved from http://www.sanger.ac.uk/science/tools/sanger-imputation-service

Saragiotto, B. T., Yamato, T. P., Junior, L. C. H., Rainbow, M. J., Davis, I. S., & Lopes, A. D. (2014). What are the main risk factors for running-related injuries? *Sports medicine, 44*(8), 1153-1163.

Sasaki, S., Ito, E., Toki, T., Maekawa, T., Kanezaki, R., Umenai, T., . . . Yamamoto, M. (2000). Cloning and expression of human B cell-specific transcription factor BACH2 mapped to chromosome 6q15. *Oncogene, 19*(33), 3739.

Saunders, C. J., Van Der Merwe, L., Posthumus, M., Cook, J., Handley, C. J., Collins, M., & September, A. V. (2013). Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *Journal of Orthopaedic Research, 31*(4), 632-637.

Scheideler, M., Elabd, C., Zaragosi, L.-E., Chiellini, C., Hackl, H., Sanchez-Cabo, F., . . . Papak, C. (2008). Comparative transcriptomics of human multipotent stem cells during adipogenesis and osteoblastogenesis. *BMC genomics, 9*(1), 340.

Scherag, A., Müller, H.-H., Dempfle, A., Hebebrand, J., & Schäfer, H. (2003). Data adaptive interim modification of sample sizes for candidate-gene association studies. *Human heredity, 56*(1-3), 56-62.

Schnohr, P., Marott, J. L., Lange, P., & Jensen, G. B. (2013). Longevity in male and female joggers: the Copenhagen City Heart Study. *American journal of epidemiology, 177*(7), 683-689.

Schnohr, P., O'Keefe, J. H., Marott, J. L., Lange, P., & Jensen, G. B. (2015). Dose of jogging and long-term mortality: the Copenhagen City Heart Study. *Journal of the American College of Cardiology, 65*(5), 411-419.

Scuteri, A., Sanna, S., Chen, W.-M., Uda, M., Albai, G., Strait, J., . . . Usala, G. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS genetics, 3*(7), e115.

Seidegård, J., & Ekström, G. (1997). The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *Environmental Health Perspectives, 105*(Suppl 4), 791.

Sensis. (2016). How Australian people and businesses are using social media. . Retrieved from https://www.sensis.com.au/asset/PDFdirectory/Sensis_Social_Media_Report_2016.PDF

September, A. V. (2009). Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *British Journal of Sports Medicine, 43*(5), 357-365.

September, A. V., Nell, E.-M., O'Connell, K., Cook, J., Handley, C. J., Merwe, L. v. d., . . . Collins, M. (2011). A pathway-based approach investigating the genes encoding interleukin-1β, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy. *British Journal of Sports Medicine, 45*(13), 1040-1047.

September, A. V., Posthumus, M., Van Der Merwe, L., Schwellnus, M., Noakes, T. D., & Collins, M. (2008). The COL12A1 and COL14A1 Genes and Achilles Tendon Injuries. *Int J Sports Med, 29*(3), 257-263.

Sharma, P., & Maffulli, N. (2005). Basic biology of tendon injury and healing. *The Surgeon: Journal Of The Royal Colleges Of Surgeons Of Edinburgh And Ireland, 3*(5), 309-316.

Shi, Z., He, F., Chen, M., Hua, L., Wang, W., Jiao, S., & Zhou, Z. (2017). DNA-binding mechanism of the Hippo pathway transcription factor TEAD4. *Oncogene, 36*(30), 4362.

Shrier, I. (1999). Stretching before exercise does not reduce the risk of local muscle injury: a critical review of the clinical and basic science literature. *Clinical Journal of Sport Medicine, 9*(4), 221-227.

Singer, A., Ben-Yehuda, O., Ben-Ezra, Z., & Zaltzman, S. (1990). Multiple identical stress fractures in monozygotic twins. Case report. *JBJS Case Connector*(3), 444-445.

Sinnatamby, C. S. (2011). *Last's anatomy: regional and applied*: Elsevier Health Sciences.

Skol, A. D., Scott, L. J., Abecasis, G. R., & Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature genetics, 38*(2), 209-213.

Skottman, H., Mikkola, M., Lundin, K., Olsson, C., Strömberg, A. M., Tuuri, T., . . . Lahesmaa, R. (2005). Gene expression signatures of seven individual human embryonic stem cell lines. *Stem cells, 23*(9), 1343-1356.

Slade, J., Teesson, W., & Burgess, P. (2009). The mental health of Australians 2: report on the 2007 National Survey of Mental Health and Wellbeing.

Sode, J., Obel, N., Hallas, J., & Lassen, A. (2007). Use of fluroquinolone and risk of Achilles tendon rupture: a population-based cohort study. *European journal of clinical pharmacology, 63*(5), 499-503.

Söderlund, A., Fischer, A., & Johansson, T. (2009). Physical activity, diet and behaviour modification in the treatment of overweight and obese adults: a systematic review. *Perspectives in public health, 129*(3), 132-142.

Spalek, K., Coynel, D., Freytag, V., Hartmann, F., Heck, A., Milnik, A., . . . Papassotiropoulos, A. (2017). A common NTRK2 variant is associated with emotional arousal and brain white-matter integrity in healthy young subjects. *Translational psychiatry, 6*(3), e758.

Spencer, C. C., Su, Z., Donnelly, P., & Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet, 5*(5), e1000477.

Stevinson, C., & Hickson, M. (2013). Exploring the public health potential of a mass community participation event. *Journal of Public Health*, fdt082.

Stevinson, C., Wiltshire, G., & Hickson, M. (2015). Facilitating participation in health-enhancing physical activity: A qualitative study of parkrun. *International journal of behavioral medicine, 22*(2), 170-177.

Stoltenberg, S. F., Lehmann, M. K., Christ, C. C., Hersrud, S. L., & Davies, G. E. (2011). Associations among types of impulsivity, substance use problems and neurexin-3 polymorphisms. *Drug and alcohol dependence, 119*(3), e31-e38.

Stults-Kolehmainen, M. A., Ciccolo, J. T., Bartholomew, J. B., Seifert, J., & Portman, R. S. (2013). Age and gender-related changes in exercise motivation among highly active individuals. *Athletic Insight, 5*(1), 45-63.

Styrkarsdottir, U., Cazier, J.-B., Kong, A., Rolfsson, O., Larsen, H., Bjarnadottir, E., . . . Christiansen, C. (2003). Linkage of osteoporosis to chromosome 20p12 and association to BMP2. *PLoS biology, 1*(3), e69.

Styrkarsdottir, U., Halldorsson, B. V., Gretarsdottir, S., Gudbjartsson, D. F., Walters, G. B., Ingvarsson, T., . . . Nguyen, T. V. (2008). Multiple genetic loci for bone mineral density and fractures. *New England Journal of Medicine, 358*(22), 2355-2365.

Sun, P. C., Uppaluri, R., Schmidt, A. P., Pashia, M. E., Quant, E. C., Sunwoo, J. B., . . . Scholnick, S. B. (2001). Transcript map of the 8p23 putative tumor suppressor region. *Genomics, 75*(1), 17-25.

Sundaravel, S., Duggan, R., Bhagat, T., Ebenezer, D. L., Liu, H., Yu, Y., . . . Liu, T.-C. (2015). Reduced DOCK4 expression leads to erythroid dysplasia in myelodysplastic syndromes. *Proceedings of the National Academy of Sciences, 112*(46), E6359-E6368.

Takeda, S., Elefteriou, F., Levasseur, R., Liu, X., Zhao, L., Parker, K. L., . . . Karsenty, G. (2002). Leptin regulates bone formation via the sympathetic nervous system. *Cell, 111*(3), 305-317.

Tanaka, T., Sohmiya, K., Kono, T., Terasaki, F., Horie, R., Ohkaru, Y., . . . Kitaura, Y. (2007). Thiamine attenuates the hypertension and metabolic abnormalities in CD36-defective SHR: uncoupling of glucose oxidation from cellular entry accompanied with enhanced protein O-GlcNAcylation in CD36 deficiency. *Molecular and cellular biochemistry, 299*(1-2), 23.

Taunton, J. E., Ryan, M. B., Clement, D. B., McKenzie, D. C., Lloyd-Smith, D. R., & Zumbo, B. D. (2002). A retrospective case-control analysis of 2002 running injuries. *British Journal of Sports Medicine, 36*(2), 95-101.

Taunton, J. E., Ryan, M. B., Clement, D. B., McKenzie, D. C., Lloyd-Smith, D. R., & Zumbo, B. D. (2003). A prospective study of running injuries: the Vancouver Sun Run "In Training" clinics. *British Journal of Sports Medicine, 37*(3), 239-244.

Tenforde, A. S., Kraus, E., & Fredericson, M. (2016). Bone stress injuries in runners. *Physical Medicine and Rehabilitation Clinics, 27*(1), 139-149.

Tenforde, A. S., Sayres, L. C., McCurdy, M. L., Collado, H., Sainani, K. L., & Fredericson, M. (2011). Overuse injuries in high school runners: lifetime prevalence and prevention strategies. *Pm r, 3*(2), 125-131; quiz 131. doi:10.1016/j.pmrj.2010.09.009

Thacker, S. B., Gilchrist, J., Stroup, D. F., & Kimsey Jr, C. D. (2004). The impact of stretching on sports injury risk: a systematic review of the literature. *Medicine & Science in Sports & Exercise, 36*(3), 371-378.

Thornton, L., Batterham, P. J., Fassnacht, D. B., Kay-Lambkin, F., Calear, A. L., & Hunt, S. (2016). Recruiting for health, medical or psychosocial research using Facebook: systematic review. *Internet Interventions, 4*, 72-81.

Thorogood, A., Mottillo, S., Shimony, A., Filion, K. B., Joseph, L., Genest, J., . . . Eisenberg, M. J. (2011). Isolated aerobic exercise and weight loss: a systematic review and meta-analysis of randomized controlled trials. *The American journal of medicine, 124*(8), 747-755.

Tilley, B. J., Cook, J. L., Docking, S. I., & Gaida, J. E. (2015). Is higher serum cholesterol associated with altered tendon structure or tendon pain? A systematic review. *British Journal of Sports Medicine*, bjsports-2015-095100.

Togari, A., & Arai, M. (2008). Pharmacological topics of bone metabolism: the physiological function of the sympathetic nervous system in modulating bone resorption. *Journal of pharmacological sciences, 106*(4), 542-546.

Torres, C. M., Siebert, M., Bock, H., Mota, S. M., Castan, J. U., Scornavacca, F., . . . Bianchin, M. M. (2017). Tyrosine receptor kinase B gene variants (NTRK2 variants) are associated with depressive disorders in temporal lobe epilepsy. *Epilepsy & Behavior, 71*, 65-72.

Torstveit, M. K., & Sundgot-Borgen, J. (2005). The female athlete triad: are elite athletes at increased risk? *Medicine & Science in Sports & Exercise, 37*(2), 184-193.

Tuan, K., Wu, S., & Sennett, B. (2004). Stress fractures in athletes: risk factors, diagnosis, and management. *Orthopedics, 27*(6), 583.

UK Biobank. (2019). Retrieved from https://www.ukbiobank.ac.uk/

Upadhyay, G., Goessling, W., North, T., Xavier, R., Zon, L., & Yajnik, V. (2008). Molecular association between β-catenin degradation complex and Rac guanine exchange factor DOCK4 is essential for Wnt/β-catenin signaling. *Oncogene, 27*(44), 5845-5855.

Usami, Y., Gunawardena, A. T., Iwamoto, M., & Enomoto-Iwamoto, M. (2016). Wnt signaling in cartilage development and diseases: lessons from animal studies. *Laboratory Investigation, 96*(2), 186-196.

Vaags, A. K., Lionel, A. C., Sato, D., Goodenberger, M., Stein, Q. P., Curran, S., . . . Senman, L. (2012). Rare deletions at the neurexin 3 locus in autism spectrum disorder. *The American Journal of Human Genetics, 90*(1), 133-141.

Välimäki, V.-V., Alfthan, H., Lehmuskallio, E., Löyttyniemi, E., Sahi, T., Suominen, H., & Välimäki, M. J. (2005). Risk factors for clinical stress fractures in male military recruits: a prospective cohort study. *Bone, 37*(2), 267-273.

van der Worp, M. P., Ten Haaf, D. S., van Cingel, R., de Wijer, A., Nijhuis-van der Sanden, M. W., & Staal, J. B. (2015). Injuries in runners; a systematic review on risk factors and sex differences. *PloS one, 10*(2), e0114937.

van Gent, R. N., Siem, D., van Middelkoop, M., van Os, A. G., Bierma-Zeinsfra, S. M. A., & Koes, B. W. (2007). Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. *British Journal of Sports Medicine, 41*(8), 469-480.

Van Mechelen, W. (1992). Running injuries. *Sports medicine, 14*(5), 320-335.

Van Mechelen, W., Hlobil, H., Kemper, H. C., Voorn, W. J., & de Jongh, H. R. (1993). Prevention of running injuries by warm-up, cool-down, and stretching exercises. *The American journal of sports medicine, 21*(5), 711-719.

Van Middelkoop, M., Kolkman, J., Van Ochten, J., Bierma-Zeinstra, S., & Koes, B. W. (2008). Risk factors for lower extremity injuries among male marathon runners. *Scandinavian Journal of Medicine & Science in Sports, 18*(6), 691-697.

Varley, I., Greeves, J. P., Sale, C., Friedman, E., Moran, D. S., Yanovich, R., . . . Stellingwerff, T. (2016). Functional polymorphisms in the P2X7 receptor gene are associated with stress fracture injury. *Purinergic signalling, 12*(1), 103-113.

Varley, I., Hughes, D. C., Greeves, J. P., Stellingwerff, T., Ranson, C., Fraser, W. D., & Sale, C. (2015). RANK/RANKL/OPG pathway: Genetic associations with stress fracture period prevalence in elite athletes. *Bone, 71*, 131-136.

Vasilyeva, N., Loktyushov, E., Bychkov, M., Shenkarev, Z., & Lyukmanova, E. (2017). Three-finger proteins from the Ly6/uPAR family: Functional diversity within one structural motif. *Biochemistry (Moscow), 82*(13), 1702-1715.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics, 90*(1), 7-24.

Vlahovich, N., Fricker, P. A., Brown, M. A., & Hughes, D. (2016). Ethics of genetic testing and research in sport: a position statement from the Australian Institute of Sport. *Br J Sports Med*, bjsports-2016-096661.

Vlahovich, N., Hughes, D. C., Griffiths, L. R., Wang, G., Pitsiladis, Y. P., Pigozzi, F., . . . Eynon, N. (2017). Genetic testing for exercise prescription and injury prevention: AIS-Athlome consortium-FIMS joint statement. *BMC genomics, 18*(8), 818.

Voegeli, G., Ramoz, N., Shekhtman, T., Courtet, P., Gorwood, P., & Kelsoe, J. R. (2016). Neurotrophin genes and antidepressant-worsening suicidal ideation: a prospective case-control study. *International Journal of Neuropsychopharmacology, 19*(11).

Walsh, A. (2017). Shin Splints. Retrieved from http://www.runfasthq.com/shin-splints/

Walter, S. D., Hart, L., McIntosh, J. M., & Sutton, J. R. (1989). The Ontario cohort study of running-related injuries. *Archives of internal medicine, 149*(11), 2561-2564.

Walther, M., Reuter, I., Leonhard, T., & Engelhardt, P. D. M. (2005). Verletzungen und überlastungsreaktionen im laufsport. *Der Orthopäde, 34*(5), 399-404.

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Canadian medical association journal, 174*(6), 801-809.

Warden, S. J., Davis, I. S., & Fredericson, M. (2014). Management and prevention of bone stress injuries in long-distance runners. *Journal of Orthopaedic & Sports Physical Therapy, 44*(10), 749-765.

Wasielewski, N. J., & Kotsko, K. M. (2007). Does eccentric exercise reduce pain and improve strength in physically active adults with symptomatic lower extremity tendinosis? A systematic review. *Journal of athletic training, 42*(3), 409.

Wen, D. Y., Puffer, J. C., & Schmalzried, T. P. (1998). Injuries in runners: a prospective study of alignment. *Clinical Journal of Sport Medicine, 8*(3), 187-194.

Willems, T. M., Witvrouw, E., De Cock, A., & De Clercq, D. (2007). Gait-related risk factors for exercise-related lower-leg pain during shod running. *Medicine & Science in Sports & Exercise, 39*(2), 330-339.

Williams Iii, D. S., McClay, I. S., & Hamill, J. (2001). Arch structure and injury patterns in runners. *Clinical biomechanics, 16*(4), 341-347.

Williams, P. T. (1997). Relationship of distance run per week to coronary heart disease risk factors in 8283 male runners The National Runners' Health Study. *Archives of internal medicine, 157*(2), 191.

Williams, R. J., Attia, E., Wickiewicz, T. L., & Hannafin, J. A. (2000). The effect of ciprofloxacin on tendon, paratenon, and capsular fibroblast metabolism. *The American journal of sports medicine, 28*(3), 364-369.

Winfield, A. C., Moore, J., Bracker, M., & Johnson, C. W. (1997). Risk factors associated with stress reactions in female Marines. *Military medicine, 162*(10), 698-702.

World Health Organisation. (2010). *Global recommendations on physical activity for health*.

World Health Organization. (2000). *Obesity: preventing and managing the global epidemic*. (9241208945). World Health Organization.

World Health Organization. (2006). *Cancer control: knowledge into action: WHO guide for effective programmes* (Vol. 2): World Health Organization.

World Health Organization. (2009). *Global health risks: mortality and burden of disease attributable to selected major risks*. In.

Wright, A. F. (2005). Genetic variation: polymorphisms and mutations. *eLS*.

Wu, X., Tu, X., Joeng, K. S., Hilton, M. J., Williams, D. A., & Long, F. (2008). Rac1 activation controls nuclear localization of β-catenin during canonical Wnt signaling. *Cell, 133*(2), 340-353.

Wu, Z., Li, Y., Li, X., Ti, D., Zhao, Y., Si, Y., . . . Han, W. (2011). LRP16 integrates into NF-κB transcriptional complex and is required for its functional activation. *PLoS One, 6*(3), e18157.

Yamashiro, T., Fukunaga, T., Yamashita, K., Kobashi, N., & Takano-Yamamoto, T. (2001). Gene and protein expression of brain-derived neurotrophic factor and TrkB in bone and cartilage. *Bone, 28*(4), 404-409.

Yanovich, R., Friedman, E., Milgrom, R., Oberman, B., Freedman, L., & Moran, D. S. (2012). Candidate gene analysis in Israeli soldiers with stress fractures. *Journal of Sports Science & Medicine, 11*(1), 147.

Zhang, K., Huentelman, M. J., Rao, F., Sun, E. I., Corneveaux, J. J., Schork, A. J., . . . Hightower, C. M. (2014). Genetic implication of a novel thiamine transporter in human hypertension. *Journal of the American College of Cardiology, 63*(15), 1542-1555.

Zhang, R., & Song, C. (2014). Loss of CSMD1 or 2 may contribute to the poor prognosis of colorectal cancer patients. *Tumor Biology, 35*(5), 4419-4423.

Zhang, W., Modén, O., & Mannervik, B. (2010). Differences among allelic variants of human glutathione transferase A2-2 in the activation of azathioprine. *Chemico-biological interactions, 186*(2), 110-117.

Zhang, X., Chen, Y., Ye, Y., Wang, J., Wang, H., Yuan, G., . . . Lin, X. (2017). Wnt signaling promotes hindgut fate commitment through regulating multi-lineage genes during hESC differentiation. *Cellular signalling, 29*, 12-22.

Zhang, Z., Yu, H., Jiang, S., Liao, J., Lu, T., Wang, L., . . . Yue, W. (2015). Evidence for association of cell adhesion molecules pathway and NLGN1 polymorphisms with schizophrenia in Chinese Han population. *PloS one, 10*(12), e0144719.

Zmuda, J. M., Yerges-Armstrong, L. M., Moffett, S. P., Klei, L., Kammerer, C. M., Roeder, K., . . . Nestlerode, C. S. (2011). Genetic analysis of vertebral trabecular bone density and cross-sectional area in older men. *Osteoporosis international, 22*(4), 1079-1090.

Zondervan, K. T., & Cardon, L. R. (2007). Designing candidate gene and genome-wide case–control association studies. *Nature protocols, 2*(10), 2492.

8. Appendices

Appendix 1 – Summary table of studies of genetic polymorphisms associated with Achilles tendinopathy

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
Abate, M., et al. (2015)	Cross-sectional study/2b	38/38	CON: 69± 2.8; AT: 69.6±3.3	CON and AT: 32 M and 6 F in each group	CON: 24.8 ± 2.3; AT:26.8 ± 3	0	Speed walking, jogging, tennis	Pain at rest or during activities in the AT region, and/or local tenderness or swelling, and/or functional limitation (ankle dorsiflexion and extension)	Diabetes itself, associated or not to overweight, could be an important contributing factor to the development of AT (p=0.004).
Abrahams, Y., et al. (2013)	Retrospective case-control study/2b	160 (81 SA+79 AUS/342 (149 SA + 93 AUS)	TEN (age of intial injury): 39.8± 14.5; CON: 37.7 ± 11.7	TEN: 73%, M 27% F CON: 50.6% M, 49.4% F	TEN 25.7 ± 3.8 kg/m2,N = 147 vs CON 24.2 ± 3.6 kg/m2,N = 330	0	long distance running, squash	As per Mokone et al. (2005)	Polymorphisms rs71746744, rs16399, rs1134170 in *COL5A1* 3'-UTR functional region are independently associated with chronic AT (p=0.008, OR=2, CI:1.2-3.3; p=0.015, OR=1.7, CI:1.1-2.7; p=0.014, OR=1.8, CI: 1.1-2.9 respectively). Polymorphism rs4919510 in *MIR608* gene is associated with chronic AT (p=0.023, OR=1.6, CI: 1.1-2.5).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
Brown K.L., et al. (2016)	Retrospective case-control study/2b	112(87 AT and 25 RUP)/ 130 CON	CON: 41.6 ± 11.6 (123); ATP: 43.9 ± 13.8 (112)	CON: 63.1% (82) M, 36.9% (48) F; ATP: 60.7% (68) M, 39.3% (44) F	CON: 25.9 ± 4.5 (123; ATP: 26.0 ± 4.0 (82)	0	not specified, physically active people	As per Mokone et al. (2005)	Two inferred allele combinations constructed from *COL5A1* rs12722, rs3196378 and rs 71746744 were found to increase risk of ATP and RUP (p=0.023, p=0.011). One of them was also significantly associated with the increased risk of TEN (p=0.011). Alternatively, the third inferred allele combination was shown to decrease risk of ATP and RUP (p=0.011, p=0.004). Inferred allele combinations were constructed from *CASP8* rs3834129 and rs1045485. One of them was significantly assocated with an increased risk of TEN (p=0.031).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
El Khoury et al. (2016)	Retrospective case-control study/2b	118(93 AT and 25 RUP)/ 130 CON	CON:41.7±11.6 (124); ATP: 43.7±13.8 (117)	CON: 62.6% (82) M, 37.4% (49) F; ATP: 60.2% (71) M, 39.8% (47) F	CON: 25.9±4.5 (123); ATP: 26.3±4.1 (86)	0	not specified, physically active people	As per Mokone et al. (2005)	*MMP3* rs679620 GG genotype was found to be overrepresented in the RUP group compared to CON (p=0.021). For the *TIMP2* rs4789932 variant was found a significant (p=0.038) difference in the genotype distribution frequency between males with ATP compared to male CON group. There was also a difference in the *TIMP2* rs4789932 genotype distribution between males with rupture compared to male controls (p=0.038).
El Khoury, L. E., et al. (2015)	Retrospective case-control study/2b	135 (60 AUS+75 SA)/239 (143 AUS+96 SA)	CON 38.2 ± 11.2 (230) TEN 40.1 ± 14.2 (129)	CON: 50.6%M 49.4% F (239); TEN: 77.4% M 22.6% F (129)	CON 24.2 ± 3.6 (235) TEN 25.7 ± 3.9 (124)	0	Running, high-impact sports	As per Mokone et al. (2005)	*FBN2* rs331079 variant was significantly associated with the risk of AT (p=0.035), but no association between the *ELN* rs2071307 variant and AT (p=0.795).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
El Khoury, L., et al. (2013)	Retrospective case-control study/2b	165 (59 AUS+114 SA)/248 (152 AUS + 96 SA)	AUS CON 38.5±11.9 (149) SA CON 37.1±10.0 (91) AUS ATP 40.3±14.1 (58) SA ATP 40.2±12.3 (107)	AUS CON: 39.7% M 60.3% F (151); SA CON 66.3% M 33.7% F (95); AUS ATP 67.8% M 32.2% F (59); SA ATP 73% M 27% F (111)	AUS CON 24.8±4.0 (150) SA CON 23.3±2.8 (93) AUS ATP 26.6±4.1 (57) SA ATP 26.0±3.9 (103)	20 Controls	long distance running, squash	As per Mokone et al. (2005)	*TIMP2* rs4789932 variant was found to be significantly associated with ATP (p=0.019). However, there was no association between this SNP and ATP. None of the other selected variants within the *ADAMTS2*, *ADAMTS5*, *ADAMTS14* and *ADAM12* genes were associated with risk of ATP in the two populations investigated (p=0.316, p=0.323, p=0.849, p=0.633 respectively).
Gaida, J. E., et al. (2009)	Cross-sectional study/2b	60/60	CON: 47±10, AT: 48±9	CON: 53% M 47% F; AT: 53% M, 47% F	CON: 25±3 AT: 25±3	0	not specified	Individuals with chronic Achilles tendon pain were diagnosed with midportion Achilles tendinopathy	AT subjects showed evidence of underlying dyslipidemia. They had higher triglyceride (TG) levels (P = 0.039), lower %HDL-C (P = 0.016), and higher TG/HDL-C ratio (P = 0.036) in comparison

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
								were included in the study.	to the matched control group. Achilles tendinopathy subjects also had elevated apolipoprotein B concentration (P = 0.017).
Gaida, J. E., et al. (2010)	Cross-sectional study/2b	25/273	CON M: 36.3 ± 11.3; CON F: 36.0 ± 10.3; AT M: 50.9 ± 10.4; AT F: 47.4 ± 10.0	CON: 40.3% M, 59.7% F; AT: 68% M, 32% F	CON M: 25.5 (3.5) CON F: 23.8 (3.2) AT M: 26.4 (3.2) AT F: 22.6 (2.6)	0	not specified	Achilles tendons were examined by the ultrasound. Each tendon was classified as having a normal or abnormal internal structure. A tendon was classified as abnormal if any of the three following conditions were met 1) one or more focal hypoechoic regions visible in both the longitudinal and	Asymptomatic ATP was more evident in men (13%) than women (5%) (p = 0.007). Men with tendon pathology were older (50.9 ± 10.4, 36.3 ± 11.3, p < 0.001), had greater WHR (0.926 ± 0.091, 0.875 ± 0.065, p = 0.039), higher android/gynoid fat mass ratio (0.616 ± 0.186, 0.519 ± 0.142, p = 0.014) and higher upper-body/lower body fat mass ratio (2.346 ± 0.630, 2.022 ± 0.467, p = 0.013). Men older than 40 years with a waist circumference >83 cm had the greatest prevalence of tendon

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
								transverse scans, 2) diffuse hypoechogenicity associated with bowing of the anterior tendon border, or 3) diffuse hypoechogenicity associated with generalised thickening of the tendon in comparison to the contralateral tendon.	pathology (33%). Women with tendon pathology were older (47.4 ± 10.0, 36.0 ± 10.3, p = 0.008), had less total fat (17196 ± 3173 g, 21626 ± 7882 g, p = 0.009), trunk fat (7367 ± 1662 g, 10087 ± 4152 g, p = 0.003) and android fat (1117 ± 324 g, 1616 ± 811 g, p = 0.005). They had lower central/peripheral fat mass ratios (0.711 ± 0.321 g, 0.922 ± 0.194 g, p = 0.004) than women with normal tendons.
Gibbon et al. (2016)	Retrospective case-control study/2b	153 (99 AUS+74 SA)/ 296 (199 AUS+97 SA)	SA cohort from Mokone et al (2005), AUS from Raleigh et al. (2009)	SA cohort from Mokone et al (2005), AUS from Raleigh	SA cohort from Mokone et al (2005), AUS from Raleigh	0	not specified, physically active people	As per Mokone et al. (2005)	*MMP3* variant rs679620, rs3025058 was inferred and found to be associated with increased risk for AT within the SA group (p= 0.012; OR: 2.88; 95%CI: 1.4 to 6.1). The 6A-G-C-G haplotype, constructed

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
				et al. (2009)	et al. (2009)				from the investigated variants, was significantly associated with reduced risk for AT (29%CON vs. 20% TEN, P = 0.037) in the AUS group.
Hay, M., et al. (2013)	Retrospective case-control study/2b	184 (78 AUS + 106 SA)/338 (177 AUS +161 SA)	AUS CON 39.4±12.3 (174) SA CON 36.4±10.8 (154) AUS TEN 40.7±14.5 (77) SA TEN 40.9±14.8 (92)	AUS CON: 40.3% M 59.7% F (157); SA CON 63.8% M 36.2 F (160); AUS TEN 71.8% M 28.2% F (78); SA TEN 67.6% M 32.4% F (105)	AUS CON 24.7±3.9 (175) SA CON 23.6±2.8 (151) AUS TEN 26.2±3.5 (75) SA TEN 24.8±3.3 (81)	0	long distance running, squash	As per Mokone et al. (2005)	None of the three investigated polymorphisms within the *COL11A1* (rs3753841 and rs1676486) and *COL11A2* (rs1799907) genes were independently associated with AT in the AUS, SA or combined cohorts.The main finding was the association of the TCT-inferred pseudohaplotype, constructed from the three polymorphisms within *COL11A1* (rs3753841 T/C and rs1676486 C/T) and *COL11A2* (rs17999079

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
									T/A), with an increased risk of AT.
Longo, U. G., et al. (2009)	Cross-sectional study/2b	85/93	CON: 52.4 ± 12.0; AT: 54.9 ± 11.8	no data	no data	0	Running, hurdle, jumping	VISA_A questionnaire was filled out by the participants in order to identify the presence of AT. If the score was less than 100, then they were examined by an orthpaedic surgeon to ascertain whether the AT diagnosis was appropriate.	There was no effect of gender on the presence of AT (p= 0.14). No significant track and field specialty effect upon the frequency of AT was found on the VISA-A questionnaire scores (p= 0.32). There were no differences in age, weight, and height between athletes who did or did not suffer from Achilles tendinopathy (p=0.20, p=0.21, p=0.46 respectively).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
Abate, M., et al. (2015)	Cross-sectional study/2b	38/39	CON: 69± 2.8; AT: 69.6±3.4	CON and AT: 32 M and 6 F in each group	CON: 24.8 ± 2.3; AT:26.8 ± 4	0	Speed walking, jogging, tennis	Pain at rest or during activities in the AT region, and/or local tenderness or swelling, and/or functional limitation (ankle dorsiflexion and extension)	Diabetes itself, associated or not to overweight, could be an important contributing factor to the development of AT (p=0.004).
Abrahams, Y., et al. (2013)	Retrospective case-control study/2b	161 (81 SA+79 AUS/342 (149 SA + 93 AUS)	TEN (age of intial injury): 39.8± 14.5; CON: 37.7 ± 11.8	TEN: 73%, M 27% F CON: 50.6% M, 49.4% F	TEN 25.7 ± 3.8 kg/m2,N = 147 vs CON 24.2 ± 3.6 kg/m2,N = 331	0	long distance running, squash	As per Mokone et al. (2005)	Polymorphisms rs71746744, rs16399, rs1134170 in *COL5A1* 3'-UTR functional region are independently associated with chronic AT (p=0.008, OR=2, CI:1.2-3.3; p=0.015, OR=1.7, CI:1.1-2.7; p=0.014, OR=1.8, CI: 1.1-2.9 respectively). Polymorphism rs4919510 in *MIR608* gene is associated with chronic AT (p=0.023, OR=1.6, CI: 1.1-2.5).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
Brown K.L., et al. (2016)	Retrospective case-control study/2b	112(87 AT and 25 RUP)/ 130 CON	CON: 41.6 ± 11.6 (123); ATP: 43.9 ± 13.8 (112)	CON: 63.1% (82) M, 36.9% (48) F; ATP: 60.7% (68) M, 39.3% (44) F	CON: 25.9 ± 4.5 (123; ATP: 26.0 ± 4.0 (82)	0	not specified, physically active people	As per Mokone et al. (2005)	Two inferred allele combinations constructed from *COL5A1* rs12722, rs3196378 and rs 71746744 were found to increase risk of ATP and RUP (p=0.023, p=0.011). One of them was also significantly associated with the increased risk of TEN (p=0.011). Alternatively, the third inferred allele combination was shown to decrease risk of ATP and RUP (p=0.011, p=0.004). Inferred allele combinations were constructed from *CASP8* rs3834129 and rs1045485. One of them was significantly assocated with an increased risk of TEN (p=0.031).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
El Khoury et al. (2016)	Retrospective case-control study/2b	118(93 AT and 25 RUP)/ 130 CON	CON:41.7±11.6 (124); ATP: 43.7±13.8 (117)	CON: 62.6% (82) M, 37.4% (49) F; ATP: 60.2% (71) M, 39.8% (47) F	CON: 25.9±4.5 (123); ATP: 26.3±4.1 (86)	0	not specified, physically active people	As per Mokone et al. (2005)	*MMP3* rs679620 GG genotype was found to be overrepresented in the RUP group compared to CON (p=0.021). For the *TIMP2* rs4789932 variant was found a significant (p=0.038) difference in the genotype distribution frequency between males with ATP compared to male CON group. There was also a difference in the *TIMP2* rs4789932 genotype distribution between males with rupture compared to male controls (p=0.038).
El Khoury, L. E., et al. (2015)	Retrospective case-control study/2b	136 (60 AUS+75 SA)/239 (143 AUS+96 SA)	CON 38.2 ± 11.2 (230) TEN 40.1 ± 14.2 (129)	CON: 50.6%M 49.4% F (239); TEN: 77.4% M 22.6% F (129)	CON 24.2 ± 3.6 (235) TEN 25.7 ± 3.9 (124)	0	Running, high-impact sports	As per Mokone et al. (2005)	*FBN2* rs331079 variant was significantly associated with the risk of AT (p=0.035), but no association between the *ELN* rs2071307 variant and AT (p=0.795).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
El Khoury, L., et al. (2013)	Retrospective case-control study/2b	166 (59 AUS+114 SA)/248 (152 AUS + 96 SA)	AUS CON 38.5±11.9 (149) SA CON 37.1±10.0 (91) AUS ATP 40.3±14.1 (58) SA ATP 40.2±12.3 (107)	AUS CON: 39.7% M 60.3% F (151); SA CON 66.3% M 33.7% F (95); AUS ATP 67.8% M 32.2% F (59); SA ATP 73% M 27% F (111)	AUS CON 24.8±4.0 (150) SA CON 23.3±2.8 (93) AUS ATP 26.6±4.1 (57) SA ATP 26.0±3.9 (103)	21 Controls	long distance running, squash	As per Mokone et al. (2005)	*TIMP2* rs4789932 variant was found to be significantly associated with ATP (p=0.016). However, there was no association between this SNP and ATP. None of the other selected variants within the *ADAMTS2*, *ADAMTS5*, *ADAMTS14* and *ADAM12* genes were associated with risk of ATP in the two populations investigated (p=0.316, p=0.323, p=0.849, p=0.633 respectively).
Gaida, J. E., et al. (2009)	Cross-sectional study/2b	60/61	CON: 47±10, AT: 48±10	CON: 53% M 47% F; AT: 53% M, 47% F	CON: 25±3 AT: 25±4	0	not specified	Individuals with chronic Achilles tendon pain were diagnosed with midportion Achilles tendinopathy	AT subjects showed evidence of underlying dyslipidemia. They had higher triglyceride (TG) levels (P = 0.039), lower %HDL-C (P = 0.016), and higher TG/HDL-C ratio (P = 0.036) in comparison

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
								were included in the study.	to the matched control group. Achilles tendinopathy subjects also had elevated apolipoprotein B concentration (P = 0.017).
Gaida, J. E., et al. (2010)	Cross-sectional study/2b	25/274	CON M: 36.3 ± 11.3; CON F: 36.0 ± 10.3; AT M: 50.9 ± 10.4; AT F: 47.4 ± 10.1	CON: 40.3% M, 59.7% F; AT: 68% M, 32% F	CON M: 25.5 (3.5) CON F: 23.8 (3.2) AT M: 26.4 (3.2) AT F: 22.6 (2.6)	0	not specified	Achilles tendons were examined by the ultrasound. Each tendon was classified as having a normal or abnormal internal structure. A tendon was classified as abnormal if any of the three following conditions were met 1) one or more focal hypoechoic regions visible in both the longitudinal and	Asymptomatic ATP was more evident in men (13%) than women (5%) (p = 0.007). Men with tendon pathology were older (50.9 ± 10.4, 36.3 ± 11.3, p < 0.001), had greater WHR (0.926 ± 0.091, 0.875 ± 0.065, p = 0.039), higher android/gynoid fat mass ratio (0.616 ± 0.186, 0.519 ± 0.142, p = 0.014) and higher upper-body/lower body fat mass ratio (2.346 ± 0.630, 2.022 ± 0.467, p = 0.013). Men older than 40 years with a waist circumference >83 cm had the greatest prevalence of tendon

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
								transverse scans, 2) diffuse hypoechogenicity associated with bowing of the anterior tendon border, or 3) diffuse hypoechogenicity associated with generalised thickening of the tendon in comparison to the contralateral tendon.	pathology (33%). Women with tendon pathology were older (47.4 ± 10.0, 36.0 ± 10.3, p = 0.008), had less total fat (17196 ± 3173 g, 21626 ± 7882 g, p = 0.009), trunk fat (7367 ± 1662 g, 10087 ± 4152 g, p = 0.003) and android fat (1117 ± 324 g, 1616 ± 811 g, p = 0.005). They had lower central/peripheral fat mass ratios (0.711 ± 0.321 g, 0.922 ± 0.194 g, p = 0.004) than women with normal tendons.
Gibbon et al. (2016)	Retrospective case-control study/2b	154 (99 AUS+74 SA)/ 296 (199 AUS+97 SA)	SA cohort from Mokone et al (2005), AUS from Raleigh et al. (2009)	SA cohort from Mokone et al (2005), AUS cohort from	SA cohort from Mokone et al (2005), AUS cohort from	0	not specified, physically active people	As per Mokone et al. (2005)	*MMP3* variant rs679620, rs3025058 was inferred and found to be associated with increased risk for AT within the SA group (p= 0.012; OR: 2.88; 95%CI: 1.4 to 6.1). The 6A-G-C-G haplotype, constructed

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
				Raleigh et al. (2009)	Raleigh et al. (2009)				from the investigated variants, was significantly associated with reduced risk for AT (29%CON vs. 20% TEN, P = 0.037) in the AUS group.
Hay, M., et al. (2013)	Retrospective case-control study/2b	185 (78 AUS + 106 SA)/338 (177 AUS +161 SA)	AUS CON 39.4±12.3 (174) SA CON 36.4±10.8 (154) AUS TEN 40.7±14.5 (77) SA TEN 40.9±14.8 (92)	AUS CON: 40.3% M 59.7% F (157); SA CON 63.8% M 36.2 F (160); AUS TEN 71.8% M 28.2% F (78); SA TEN 67.6% M 32.4% F (105)	AUS CON 24.7±3.9 (175) SA CON 23.6±2.8 (151) AUS TEN 26.2±3.5 (75) SA TEN 24.8±3.3 (81)	0	long distance running, squash	As per Mokone et al. (2005)	None of the three investigated polymorphisms within the *COL11A1* (rs3753841 and rs1676486) and *COL11A2* (rs1799907) genes were independently associated with AT in the AUS, SA or combined cohorts. The main finding was the association of the TCT-inferred pseudohaplotype, constructed from the three polymorphisms within *COL11A1* (rs3753841 T/C and rs1676486 C/T) and *COL11A2* (rs17999079

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
									T/A), with an increased risk of AT (p<0.05).
Longo, U. G., et al. (2009)	Cross-sectional study/2b	85/94	CON: 52.4 ± 12.0; AT: 54.9 ± 11.9	no data	no data	0	Running, hurdle, jumping	VISA_A questionnaire was filled out by the participants in order to identify the presence of AT. If the score was less than 100, then they were examined by an orthopaedic surgeon to ascertain	There was no effect of gender on the presence of AT (p= 0.14). No significant track and field specialty effect upon the frequency of AT was found on the VISA-A questionnaire scores (p= 0.32). There were no differences in age, weight, and height between athletes who did or did not suffer from Achilles

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included/ analyzed (%)	Physical activity	Injury Definition	Identified factors
								whether the AT diagnosis was appropriate.	tendinopathy (p=0.20, p=0.21, p=0.46 respectively).

Appendix 2 – Summary table of studies of genetic polymorphisms associated with bone stress injuries

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included / analyzed (%)	Physical activity	Analysed characteristics	Identified factors
Ferrari et al., (1998)	Cross-sectional	independent analyses of 197 girls under 18 years of age, and 172 women of 18-56 years of age	N/A	100% F	N/A	0	N/A	Bone mineral density (BMD)	BB genotype in *VDR* gene was associated with lower BMD in girls under 18 years of age (*p*=0.03)
Yanovich et al., 2012	Cross-sectional	Military personell: 182 cases with stress fracture (SF)/ 203 uninjured controls	SF: 20.1±1.7 ; CON: 20.2±1.3	SF: 162 M and 41 F; CON: 165 M and 17 F	SF: 23±0.3; CON: 23.2±0.2	0	Were physically active prior recruitment: SF: 81.2%; CON: 78.9%	Lower limb stress fracture	CCAGGCAC (8 SNPs) haplotype of the VDR gene was associated with 12.22-fold increased risk of SF (95%CI 1.45-102.7, p = 0.022). CGTTCTCCGA (10 SNPs) haplotype of the CALCR gene was associated with 1.93-fold increased risk of SF (95%CI 1.11-3.50, p = 0.00255).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included / analyzed (%)	Physical activity	Analysed characteristics	Identified factors
Zmuda et al., 2011	Cross-sectional	2018 0lder men from Osteoporotic Fractures in Men Study	74±5.9	100% M	27.6±3.9	0	N/A	volumetric BMD (vBMD)	Rs1877632 in *SOST* gene, rs1801197 in *CALCR* gene, *rs1888057* in *KL* gene were associated with phenotypic variance in vertebral trabecular vBMD (*p*=0.0001, *p*=0.0037, *p*=0.0035 respectively). SNPs in TGFBR3, MEPE, PTN, FGFR2, LEP, CSF1R and GNRH2 were also associated with BMD in this study.
Styrkarsdottir et al., 2003	Cross-sectional	Icelandic population: 957 osteoporotic fractures cases/ 710 randomly selected controls form the	N/A	CASE: 176 M and 781 F; CON: 242 M and 468 F	N/A	0	N/A	BMD and osteoporotic fractures	Ser37Ala variant of *BMP2* gene was associated with osteoporotic fractures and low BMD (*p*=0.011; *p*=00068 respectively).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included / analyzed (%)	Physical activity	Analysed characteristics	Identified factors
		general population							
Varley et al., 2015	Cross-sectional	518 elite athletes: 125 stress fracture cases/ 376 unaffected controls	SF: 27.7±7.5 ; CON: 24.4±5.4	SF: 98 M and 27 F; CON: 335 M and 41 F	SF: 23.2±2.7 ; CON: 23.7±2.2	3.3% (n=17)	Sports: football, cricket, track and field, rowing, boxing, tennis, hockey, gymnastics	Radiologically confirmed stress fractures	SNPs rs3018362 (*RANK*), rs1021188 (*RANKL*)within the RANK/RANKL/OPG signalling pathway were associated with stress fracture injuries and rs4355801(*OPG*) was associated with increased risk of multiple stress injuries (allele G, *p*=0.008, OR=1.41, CI: 1.05-1.91; genotype AA, *p*=0.024, OR=2.93, CI: 1.18-7.28; GA or GG genotype, p=0.042, OR=2.05, CI: 1.01-4.17 respectively).

Author	Study design/ level of evidence	Case/Control Numbers	Age (yr) (mean ± sd)	Sex (%)	BMI (kg/m2)	Not included / analyzed (%)	Physical activity	Analysed characteristics	Identified factors
Varley et al., 2016	Cross-sectional	Military personell (MP): 43 stress fracture cases/ 167 unaffected controls. Elite athletes (EA): 125 stress fracture cases/ 376 unaffected controls	MP: SF: 20.3±1.6 ; CON: 18.9±0.5 . EA: SF: 27.7±7.5 ; CON: 24.4±5.4	MP: SF: 41 M and 2 F; CON: 157 M and 10 F. EA: SF: 98 M and 27 F; CON: 335 M and 41 F	MP: SF: 22.6±1.8 ; CON: 23.2±2.6 . EA: SF: 23.2±2.7 ; CON: 23.7±2.2	0	MP: Military service training. EA: sports: football, cricket, track and field, rowing, boxing, tennis, hockey, gymnastics	Radiologically confirmed stress fractures	In both MP and EA cohorts, rs3751143 in *P2X7R* gene were associated with bone stress fractures ($p<0.05$). Additionally, rs1718119 in P2X7R gene was associated with stress fractures in MP cohort ($p<0.01$).

Appendix 3 – Ethics application - Removed.

Appendix 4 – Ethics approval - Removed

Appendix 5 – Online Questionnaire

The genetics of exercise-induced injuries in tendon and bone

Page exit logic: Disqualify if user does not accept conditions
IF: Question "Do you accept to participate in this research study?" is one of the following answers ("**I DO NOT ACCEPT** the conditions above") **THEN:** Disqualify and display:
You must accept the conditions of the survey to participate in this study. Thank you for your time.

Principal Investigator
Dr David Hughes (Department of Sports Medicine, Australian Institute of Sport)
Co-Investigators
Dr Nicole Vlahovich, Dr Stacey Compton, Maria Kozlovskaia and Dr Renae Domaschenz (Department of Sports Medicine, Australian Institute of Sport), A/Prof Bon Gray, A/Prof Lotti Tajouri, A/Prof Justin Keogh, A/Prof Mike Climstein, Rebecca Grealy, A/Prof Kevin Ashton and Professor Nuala Byrne (Bond University), Professor Matthew Brown and Dr Paul Leo (University of Queensland) and Professor Maria Fiatarone Singh, Dr Yorgi Mavros, Guy Wilson and Jacinda Meiklejohn (University of Sydney).
Ethics Protocol Number
RO1688B

Welcome!

- You are invited to take part in this research project if you are:
  - **A recreational runner**
  - **Aged 18 and over**
  - **Run over 15 km via 2-5 sessions per week**
- Please read the Participant Information Form carefully as this will tell you about the research project and explain what is involved. This will help you decide if you want to continue and take part.
- Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not to take part, you might want to talk about it with a relative, friend or healthcare worker.

Participation

- Participation in this study is entirely voluntary.
- You're not obliged to participate and if you do, you can withdraw at any time without penalty or prejudice.

- To participate, we would like you to complete this online questionnaire, providing details of your medical history, injury history and running habits.
- This survey should take no more than 30 minutes to complete.
- You are able to exit the survey and complete at a later date using the link at the top of the page.
- Your participation, personal details and results will be strictly confidential and only the principal researchers above will have access to this information

By ticking the '**I ACCEPT**' option below you are telling us that you:

- understand what you have read;
- consent to take part in the research project;
- consent to participate in the research processes that are described;
- consent to the use of your personal and health information as described;
- understand that you are free to not answer specific items or questions in interviews or questionnaires;
- understand that any data or answers to questions will remain confidential with regard to your identity;
- certify to the best of your knowledge and belief, you have no physical or mental illness or weakness that would increase the risk of participating in this project;
- are participating in this project of your own free will and have not been coerced in any way to participate.

Do you accept to participate in this research study? *

- ○ **I ACCEPT** the conditions above
- ○ **I DO NOT ACCEPT** the conditions above

Further information about the project is available in the Participant Information Form

- ☐ I wish to see more information about the project:

Participant Information

Page entry logic:
This page will show when: Question "Further information about the project is available in the Participant Information Form" is one of the following answers ("I wish to see more information about the project:")

What is the purpose of this study?
Identifying new genes or new mutations in known genes will help researchers to better understand the relationship between lifestyle, health status and genetic profile and we hope that this information will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases.

This research is being conducted by a collaboration between Bond University, the Australian Institute of Sport, the University of Queensland and the University of Sydney.

What does participation in the research project involve?
Participation in this project involves the completion of an online questionnaire and the agreement to be followed up by a researcher at the Australian Institute of Sport for the provision of a saliva sample for genetic analysis. Not all participants will be followed up to provide a genetic sample

What will happen to my questionnaire submission?
Your information will be stored in password protected databases. Information about you including medical information that you provide will be treated in such a way that you cannot be identified in publications, except with your permission. In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except with your permission.

Any information obtained for the purpose of this research project that can identify you will be treated as confidential and securely stored. It will be disclosed only with your permission, or as permitted by law.

What are the possible benefits to participating?
We cannot guarantee that you will receive any benefits from this research, but your participation in the study may help doctors to better understand the relationship between lifestyle, health status and genetic profile with the hope that this will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases.

What are the possible risks to participating in the study?
All medical information is stored in password protected databases. It is possible, though very unlikely, that someone could get access to this database without permission. During the research project, new information about the risks and benefits of the project may become known

to the researchers. If this occurs, you will be told about this new information and the researcher will discuss whether this new information affects you.

If you become upset or distressed as a result of your participation in the research, the researcher is able to arrange for counselling or other appropriate support. Any counselling or support will be provided by staff who are not members of the research team. In addition, you may prefer to suspend or end your participation in the research if distress occurs.

Can anyone participate in this study?
As long as you meet the criteria specified earlier, you are eligible to take part. The research team have allocated the research funding to areas in which the technology is most likely to find new genes

Do I have to take part in this research project?
Participation in any research project is voluntary. If you do not wish to take part then you don't have to. If you decide to take part and later change your mind, you are free to withdraw at any stage. All information that you have provided can be destroyed at any time. You can withdraw your consent to participate in this research project by emailing the Principal Investigator at injurystudy@ausport.gov.au

Before you make a decision to participate in any follow up studies, a member of the research team will contact you so that you can ask any questions you have about the project. You can ask for any information that you want.

How will I be informed of results from of this research project?
In accordance with relevant Australian privacy and other relevant laws, you have the right to access the information collected and stored by the researchers about you. You also have the right to request that any information, with which you disagree, be corrected. Please contact one of the researchers named at the end of this document if you would like to access your information.

Is this research project approved?
This project will be carried out according to the *National Statement on Ethical Conduct in Research Involving Humans* (2007) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies.

The ethical aspects of this research project have been approved by the Bond University Human Research Ethics Committee, protocol number RO1688B.

Will I get paid to participate in this study?
You will not be paid for participating in this study. Participation by completion of the online survey will enable the participant to be entered into a draw to win a $50 voucher to a sporting

goods store or similar. This draw will be done randomly allowing for a 1 in 200 chance of 'winning' the voucher.

Who can I contact if I have any questions or problems in relation to this study?
If you wish to discuss further the experimental procedure or have any questions, please do not hesitate to contact Dr Nicole Vlahovich or Dr Renae Domaschenz phone (02) 6214 1578 or email injurystudy@ausport.gov.au

If you have any concerns with respect to the conduct of this study, you may contact the Secretary of the Bond University Human Research Ethics Committee Dr Lisa Marlow on (07) 5595 4194 or by email lmarlow@bond.edu.au

Your Personal Details

Page exit logic: Year of birth disqualification
IF: Question "What year were you born?" is greater than "1997" **THEN:** Disqualify and display: We're sorry, you must be aged 18 or older to participate in this study. Thank you for your time.

1. Please enter your contact information here:

First Name *

Last Name *

Street Address

City

State

ACT
NSW
NT
QLD
SA
TAS
VIC
WA

Postcode

Email Address *

Phone Number

2. Please enter your personal details here

What year were you born? *

What month were you born?

January
February
March
April
May
June
July
August
September
October
November

November
December

What date of the month were you born?

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Sex *

Male
Female

Weight (kgs) *

Height (cm) *

I give permission, if I am eligible, to be contacted in the future for related research *

- Yes
- No

I give permission, if I am eligible, to be contacted in the future to provide a saliva sample for genetic related analysis *

- Yes
- No

Your Ethnic Background

3. What is your country of birth?

- Australia
- Other

4. What is your country of citizenship?

- Australia
- Other

5. What is the ethnic background of your biological grandparents?

Maternal Grandmother	Caucasian European Mediterranean Asian African Polynesian Indigenous Australian or TSI Unknown Other
Maternal Grandfather	Caucasian European Mediterranean Asian African Polynesian Indigenous Australian or TSI Unknown Other
Paternal Grandmother	Caucasian European Mediterranean Asian African Polynesian Indigenous Australian or TSI Unknown Other
Paternal Grandfather	Caucasian European Mediterranean Asian African Polynesian Indigenous Australian or TSI Unknown Other

Your Running Habits

6. Which is your dominate leg* (used for kicking a ball)? *If you are unsure please choose according to whether you are left or right handed

- Left
- Right
- Ambidextrous

7. How many years have you been running on a regular* basis? *Regular is defined as at least weekly.

	< 1	1	2	3	4	5	6	7	8	9	10 +
Years	< 1	1	2	3	4	5	6	7	8	9	10 +

8. On average, how many km per week would you run?

	<15	15-20	20-30	30-40	40-50	50-60	60 +
Km	<15	15-20	20-30	30-40	40-50	50-60	60 +

9. Do you run every day?

- Yes
- No

10. For an average week, please indicate how many running sessions you participate in.

	1	2-3	4-5	6+
sessions/week	1	2-3	4-5	6+

11. What type of terrain is the majority of your running performed on?

- Bitumen
- Cement
- Hard dirt or gravel
- Sand
- Grass
- Synthetic
- Treadmill

12. What is your current race pace in min/km?

	> 7	6-7	5-6	4-5	< 4
min/km	> 7	6-7	5-6	4-5	< 4

13. Do you spend time stretching in association with your running session?

- Yes
- No

LOGIC Hidden unless: Question "Do you spend time stretching in association with your running session?" #13 is one of the following answers ("Yes")

14. If Yes, when do you stretch?

- Before running
- After running
- Both before and after running

LOGIC Show/hide trigger exists.

15. While running do you wear orthotics?

- ○ Yes
- ○ No

LOGIC Hidden unless: Question "While running do you wear orthotics?" #15 is one of the following answers ("Yes") Dynamically shown if "While running do you wear orthotics?" = Yes

16. If Yes, which foot?

- ○ Left foot
- ○ Right foot
- ○ Both feet

LOGIC Hidden unless: Question "While running do you wear orthotics?" #15 is one of the following answers ("Yes") Dynamically shown if "While running do you wear orthotics?" = Yes

17. If Yes, are they custom made?

- ○ Yes
- ○ No

18. What proportion of your running is: (please ensure that total of all entries equals 100%)
*minimalist means with no support or cushioning, e.g. aqua shoes, vibram five fingers (does not include Nike free, etc.)

0 with standard running shoes?

0 with minimalist running shoes?

0 barefoot?

0 out of 100 Total

LOGIC Show/hide trigger exists.
19. In the last two years have you participated in any other sports or intentional exercise on a regular basis (for example weekly during at least one season)?

- Yes
- No

LOGIC Hidden unless: Question "In the last two years have you participated in any other sports or intentional exercise on a regular basis (for example weekly during at least one season)?" #19 is one of the following answers ("Yes") Dynamically shown if "In the last two years have you participated in any other sports or intentional exercise on a regular basis (for example weekly during at least one season)?" = Yes
20. If yes, what sports? (please list all)

Running Related Injuries

LOGIC Show/hide trigger exists.
21. In the **last 2 years** have you had any injuries of the lower limbs, which have forced you to discontinue running for a period of 2 weeks or more? *

- Yes
- No

LOGIC Hidden unless: Question "In the **last 2 years** have you had any injuries of the lower limbs, which have forced you to discontinue running for a period of 2 weeks or more?" #21 is one of the following answers ("Yes") Dynamically shown if "In the **last 2 years** have you had any injuries of the lower limbs, which have forced you to discontinue running for a period of 2 weeks or more?" = Yes

22. If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?

- 1
- 2
- 3
- 4 +

Injury 1 - Your Most Recent Injury

Page entry logic:
This page will show when: (Question "In the **last 2 years** have you had any injuries of the lower limbs, which have forced you to discontinue running for a period of 2 weeks or more?" #21 is one of the following answers ("Yes") AND Question "If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?" #22 is one of the following answers ("1","2","3","4 +"))

Please answer the following questions in relation to **your most recent lower limb injury.**

23. You indicated that you have been diagnosed with a lower limb injury within the past 2 years How did this injury occur?

- while running
- while walking
- due to a fall
- during participation in another sport
- other

24. Was this injury diagnosed by a professional:

	Yes	No
Doctor?	○ Yes	○ No
Physical Therapist?	○ Yes	○ No

25. Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?

- ○ Yes
- ○ No

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #25 is one of the following answers ("Yes")

26. If imaging, what type of diagnostic imaging?

- ○ x-ray
- ○ ultrasound
- ○ bone scan
- ○ CT scan
- ○ MRI

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #25 is one of the following answers ("Yes")

27. Do you have a copy of the report of the imaging findings?

- ○ Yes
- ○ No

28. Was the injury an Achilles tendon injury?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #28 is one of the following answers ("Yes")
29. If Yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #28 is one of the following answers ("No")
30. Was the injury a bone stress injury below the knee?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury a bone stress injury below the knee?" #30 is one of the following answers ("Yes")
31. If yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: (Question "Was the injury an Achilles tendon injury?" #28 is one of the following answers ("No") AND Question "Was the injury a bone stress injury below the knee?" #30 is one of the following answers ("No"))

32. Was your injury a different injury (other than an Achilles tendon or bone stress injury)?

- ○ Yes
- ○ No

LOGIC Hidden unless: Question "Was your injury a different injury (other than an Achilles tendon or bone stress injury)?" #32 is one of the following answers ("Yes")

33. What type of injury was it?

34. Please provide details of the signs and symptoms of the injury (tick all that apply):

- ☐ bleeding
- ☐ laceration
- ☐ swelling
- ☐ bruising
- ☐ tenderness to touch
- ☐ pain at rest
- ☐ pain on movement
- ☐ instability of joint
- ☐ weakness
- ☐ numbness
- ☐ loss of sensation
- ☐ other

35. When is/was your pain at its worst?

- pain during warm up
- pain during exercise
- pain after exercise
- pain at night
- unable to exercise due to pain

36. As a result of this injury, how many weeks did you discontinue running?

	1	2	3	4	5	6	6 +
weeks	1	2	3	4	5	6	6 +

37. How and when did your symptoms start?

- sudden onset after injury
- gradual onset of pain
- pain during running
- other

38. What type of treatment did you have in association with this injury?

- Rest only
- Other treatment

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #38 is one of the following answers ("Other treatment")

39. What type of treatment did you have in association with this injury?

- [ ] medication
- [ ] physical therapy
- [ ] surgery
- [ ] bracing/taping

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #39 is one of the following answers ("medication")

40. If medication please provide details

41. Had you made any changes to your regular training program just prior to the onset of injury? *(for example, increase in training load, change in footwear, change in terrain)*

- ○ Yes
- ○ No

LOGIC Hidden unless: Question "Had you made any changes to your regular training program just prior to the onset of injury? *(for example, increase in training load, change in footwear, change in terrain)*" #41 is one of the following answers ("Yes")

42. If yes, please provide details:

43. Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)

- Yes
- No

LOGIC Hidden unless: Question "Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)" #43 is one of the following answers ("Yes")
44. If yes, please provide details:

45. Please upload any associated medical reports in relation to this injury.
You may upload a scanned report/image or a smartphone picture of the report/image (*accepted file types include png, jpg, doc, xls, docx, xlsx, pdf, txt maximum file size 1 MB*).

Browse... Choose File No file selected Upload

46. If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?

- Yes
- No
- N/A

LOGIC Hidden unless: Question "If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?" #46 is one of the following answers ("Yes")

47. Please provide name and contact details of the medical provider who holds these records.

We will send you a permission slip to sign and a stamped envelope addressed to this medical provider.
This is necessary for the release of your records to us for this survey.

Name of medical provider:	
Address of medical provider:	
Phone contact for medical provider:	

Injury 2 - Your Second Most Recent Injury

Page entry logic:
This page will show when: (Question "In the **last 2 years** have you had any injuries of the lower limbs, which have forced you to discontinue running for a period of 2 weeks or more?" #21 is one of the following answers ("Yes") AND Question "If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?" #22 is one of the following answers ("2","3","4 +"))

Please answer the following questions in relation to **your second most recent lower limb injury.**

48. You indicated that you have been diagnosed with more than 1 lower limb injury within the past 2 years
How did this injury occur?

- while running
- while walking
- due to a fall
- during participation in another sport
- other

49. Was this injury diagnosed by a professional:

	Yes	No
Doctor?	Yes	No
Physical Therapist?	Yes	No

50. Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?

- Yes
- No

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #50 is one of the following answers ("Yes")
51. If imaging, what type of diagnostic imaging?

- x-ray
- ultrasound
- bone scan
- CT scan
- MRI

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #50 is one of the following answers ("Yes")
52. Do you have a copy of the report of the imaging findings?

- Yes
- No

53. Was the injury an Achilles tendon injury?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #53 is one of the following answers ("Yes")
54. If Yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #53 is one of the following answers ("No")

55. Was the injury a bone stress injury below the knee?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury a bone stress injury below the knee?" #55 is one of the following answers ("Yes")

56. If yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: (Question "Was the injury an Achilles tendon injury?" #53 is one of the following answers ("No") AND Question "Was the injury a bone stress injury below the knee?" #55 is one of the following answers ("No"))

57. Was your injury a different injury (other than an Achilles tendon or bone stress injury)?

- Yes
- No

LOGIC Hidden unless: Question "Was your injury a different injury (other than an Achilles tendon or bone stress injury)?" #57 is one of the following answers ("Yes")

58. What type of injury was it?

59. Please provide details of the signs and symptoms of the injury (tick all that apply):

- [ ] bleeding
- [ ] laceration
- [ ] swelling
- [ ] bruising
- [ ] tenderness to touch
- [ ] pain at rest
- [ ] pain on movement
- [ ] instability of joint
- [ ] weakness
- [ ] numbness
- [ ] loss of sensation
- [ ] other

60. When is/was your pain at its worst?

- ○ pain during warm up
- ○ pain during exercise
- ○ pain after exercise
- ○ pain at night
- ○ unable to exercise due to pain

61. As a result of this injury, how many weeks did you discontinue running?

	1	2	3	4	5	6	6 +
weeks	○1	○2	○3	○4	○5	○6	○6 +

62. How and when did your symptoms start?

- ○ sudden onset after injury
- ○ gradual onset of pain
- ○ pain during running
- ○ other

63. What type of treatment did you have in association with this injury?

- ○ Rest only
- ○ Other treatment

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #63 is one of the following answers ("Other treatment")

64. What type of treatment did you have in association with this injury?

- ☐ medication
- ☐ physical therapy
- ☐ surgery
- ☐ bracing/taping

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #64 is one of the following answers ("medication")

65. If medication please provide details

66. Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)

- Yes
- No

LOGIC Hidden unless: Question "Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)" #66 is one of the following answers ("Yes")

67. If yes, please provide details:

68. Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)

- Yes
- No

LOGIC Hidden unless: Question "Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)" #68 is one of the following answers ("Yes")
69. If yes, please provide details:

70. Please upload any associated medical reports in relation to this injury.
You may upload a scanned report/image or a smartphone picture of the report/image (*accepted file types include png, jpg, doc, xls, docx, xlsx, pdf, txt maximum file size 1 MB*).

Browse... Choose File No file selected Upload

71. If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?

- Yes
- No
- N/A

LOGIC Hidden unless: Question "If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?" #71 is one of the following answers ("Yes")

72. Please provide name and contact details of the medical provider who holds these records.

We will send you a permission slip to sign and a stamped envelope addressed to this medical provider.
This is necessary for the release of your records to us for this survey.

Name of medical provider:

Address of medical provider:

Phone contact for medical provider:

Re-occuring Injury

Page entry logic:
This page will show when: Question "If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?" #22 is one of the following answers ("2","3","4 +")

LOGIC Show/hide trigger exists.

73. You have indicated that you have had more than one lower limb injury.

Was the diagnosis of the 2nd injury the same as the first injury?

- Yes
- No

LOGIC Hidden unless: Question "You have indicated that you have had more than one lower limb injury.

Was the diagnosis of the 2nd injury the same as the first injury?" #73 is one of the following answers ("Yes") Dynamically shown if "You have indicated that you have had more than one lower limb injury.

Was the diagnosis of the 2nd injury the same as the first injury?" = Yes

74. If yes, was it the same foot/leg?

- Yes
- No

75. How many days, weeks, months passed between injuries?

Days

Weeks

Months

76. Were the symptoms of the second injury the same, better or worse than the first injury?

- Same
- Better
- Worse

77. Was the treatment the same for both injuries?

- Yes
- No

78. Was the recovery longer, shorter or similar in duration?

- Longer
- Shorter
- Similar

LOGIC Show/hide trigger exists.
79. Do you have a family history of exercise-related or other lower extremity injury?

- Yes
- No

LOGIC Hidden unless: Question "Do you have a family history of exercise-related or other lower extremity injury?" #79 is one of the following answers ("Yes") Dynamically shown if "Do you have a family history of exercise-related or other lower extremity injury?" = Yes
80. If yes, was it:

	Yes	No
bone stress?	Yes	No
Achilles tendinopathy?	Yes	No
another injury?	Yes	No

LOGIC Hidden unless: Question "Do you have a family history of exercise-related or other lower extremity injury?" #79 is one of the following answers ("Yes") Dynamically shown if "Do you have a family history of exercise-related or other lower extremity injury?" = Yes

81. What family member did it occur in?

- check as many as relevant

- [ ] Maternal Grandmother
- [ ] Maternal Grandfather
- [ ] Paternal Grandmother
- [ ] Paternal Grandfather
- [ ] Mother
- [ ] Father
- [ ] Brother
- [ ] Sister
- [ ] Maternal Aunt
- [ ] Paternal Aunt
- [ ] Maternal Uncle
- [ ] Paternal Uncle
- [ ] Maternal 1st Cousin
- [ ] Paternal 1st Cousin

Injury 3 - Your Third Most Recent Injury

Page entry logic:
This page will show when: Question "If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?" #22 is one of the following answers ("3","4 +")

Please answer the following questions in relation to **your third most recent lower limb injury.**

82. You indicated that you have been diagnosed with more than 1 lower limb injury within the past 2 years
How did this injury occur?

- while running
- while walking
- due to a fall
- during participation in another sport
- other

83. Was this injury diagnosed by a professional:

	Yes	No
Doctor?	Yes	No
Physical Therapist?	Yes	No

84. Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?

- Yes
- No

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #84 is one of the following answers ("Yes")
85. If imaging, what type of diagnostic imaging?

- x-ray
- ultrasound
- bone scan
- CT scan
- MRI

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #84 is one of the following answers ("Yes")
86. Do you have a copy of the report of the imaging findings?

- Yes
- No

87. Was the injury an Achilles tendon injury?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #87 is one of the following answers ("Yes")
88. If Yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #87 is one of the following answers ("No")

89. Was the injury a bone stress injury below the knee?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury a bone stress injury below the knee?" #89 is one of the following answers ("Yes")

90. If yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: (Question "Was the injury an Achilles tendon injury?" #87 is one of the following answers ("No") AND Question "Was the injury a bone stress injury below the knee?" #89 is one of the following answers ("No"))

91. Was your injury a different injury (other than an Achilles tendon or bone stress injury)?

- Yes
- No

LOGIC Hidden unless: Question "Was your injury a different injury (other than an Achilles tendon or bone stress injury)?" #91 is one of the following answers ("Yes")

92. What type of injury was it?

93. Please provide details of the signs and symptoms of the injury (tick all that apply):

- [ ] bleeding
- [ ] laceration
- [ ] swelling
- [ ] bruising
- [ ] tenderness to touch
- [ ] pain at rest
- [ ] pain on movement
- [ ] instability of joint
- [ ] weakness
- [ ] numbness
- [ ] loss of sensation
- [ ] other

94. When is/was your pain at its worst?

- ○ pain during warm up
- ○ pain during exercise
- ○ pain after exercise
- ○ pain at night
- ○ unable to exercise due to pain

95. As a result of this injury, how many weeks did you discontinue running?

	1	2	3	4	5	6	6 +
weeks	○1	○2	○3	○4	○5	○6	○6 +

96. How and when did your symptoms start?

- ○ sudden onset after injury
- ○ gradual onset of pain
- ○ pain during running
- ○ other

97. What type of treatment did you have in association with this injury?

- ○ Rest only
- ○ Other treatment

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #97 is one of the following answers ("Other treatment")

98. What type of treatment did you have in association with this injury?

- ☐ medication
- ☐ physical therapy
- ☐ surgery
- ☐ bracing/taping

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #98 is one of the following answers ("medication")

99. If medication please provide details

100. Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)

- Yes
- No

LOGIC Hidden unless: Question "Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)" #100 is one of the following answers ("Yes")

101. If yes, please provide details:

102. Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)

- Yes
- No

LOGIC Hidden unless: Question "Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)" #102 is one of the following answers ("Yes")
103. If yes, please provide details:

104. Please upload any associated medical reports in relation to this injury.
You may upload a scanned report/image or a smartphone picture of the report/image (*accepted file types include png, jpg, doc, xls, docx, xlsx, pdf, txt maximum file size 1 MB*).

Browse... Choose File No file selected Upload

105. If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?

- Yes
- No
- N/A

LOGIC Hidden unless: Question "If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?" #105 is one of the following answers ("Yes")

106. Please provide name and contact details of the medical provider who holds these records.

We will send you a permission slip to sign and a stamped envelope addressed to this medical provider.
This is necessary for the release of your records to us for this survey.

Name of medical provider:	
Address of medical provider:	
Phone contact for medical provider:	

Injury 4 - Your Fourth Most Recent Injury

Page entry logic:
This page will show when: Question "If Yes, How many lower limb injuries have you been diagnosed with in the past 2 years?" #22 is one of the following answers ("4 +")

Please answer the following questions in relation to **your fourth most recent lower limb injury.**

107. You indicated that you have been diagnosed with more than 1 lower limb injury within the past 2 years
How did this injury occur?

- while running
- while walking
- due to a fall
- during participation in another sport
- other

108. Was this injury diagnosed by a professional:

	Yes	No
Doctor?	Yes	No
Physical Therapist?	Yes	No

109. Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?

- Yes
- No

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #109 is one of the following answers ("Yes")
110. If imaging, what type of diagnostic imaging?

- x-ray
- ultrasound
- bone scan
- CT scan
- MRI

LOGIC Hidden unless: Question "Was this injury diagnosed by imaging (x-ray/ultrasound/bone scan/CT scan/MRI)?" #109 is one of the following answers ("Yes")
111. Do you have a copy of the report of the imaging findings?

- Yes
- No

112. Was the injury an Achilles tendon injury?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #112 is one of the following answers ("Yes")
113. If Yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: Question "Was the injury an Achilles tendon injury?" #112 is one of the following answers ("No")

114. Was the injury a bone stress injury below the knee?

- Yes
- No

LOGIC Hidden unless: Question "Was the injury a bone stress injury below the knee?" #114 is one of the following answers ("Yes")

115. If yes, which leg?

- Right leg
- Left leg
- Both legs

LOGIC Hidden unless: (Question "Was the injury an Achilles tendon injury?" #112 is one of the following answers ("No") AND Question "Was the injury a bone stress injury below the knee?" #114 is one of the following answers ("No"))

116. Was your injury a different injury (other than an Achilles tendon or bone stress injury)?

- Yes
- No

LOGIC Hidden unless: Question "Was your injury a different injury (other than an Achilles tendon or bone stress injury)?" #116 is one of the following answers ("Yes")

117. What type of injury was it?

118. Please provide details of the signs and symptoms of the injury (tick all that apply):

- [ ] bleeding
- [ ] laceration
- [ ] swelling
- [ ] bruising
- [ ] tenderness to touch
- [ ] pain at rest
- [ ] pain on movement
- [ ] instability of joint
- [ ] weakness
- [ ] numbness
- [ ] loss of sensation
- [ ] other

119. When is/was your pain at its worst?

- ○ pain during warm up
- ○ pain during exercise
- ○ pain after exercise
- ○ pain at night
- ○ unable to exercise due to pain

120. As a result of this injury, how many weeks did you discontinue running?

	1	2	3	4	5	6	6 +
weeks	○1	○2	○3	○4	○5	○6	○6 +

121. How and when did your symptoms start?

- ○ sudden onset after injury
- ○ gradual onset of pain
- ○ pain during running
- ○ other

122. What type of treatment did you have in association with this injury?

- ○ Rest only
- ○ Other treatment

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #122 is one of the following answers ("Other treatment")

123. What type of treatment did you have in association with this injury?

- ☐ medication
- ☐ physical therapy
- ☐ surgery
- ☐ bracing/taping

LOGIC Hidden unless: Question "What type of treatment did you have in association with this injury?" #123 is one of the following answers ("medication")

124. If medication please provide details

125. Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)

- Yes
- No

LOGIC Hidden unless: Question "Had you made any changes to your regular training program just prior to the onset of injury?
(for example, increase in training load, change in footwear, change in terrain)" #125 is one of the following answers ("Yes")

126. If yes, please provide details:

127. Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)

- Yes
- No

LOGIC Hidden unless: Question "Had you begun or changed your participation in any new exercise other than running prior to the injury?
(for example basketball, touch football, tennis, another sport etc.)" #127 is one of the following answers ("Yes")
128. If yes, please provide details:

129. Please upload any associated medical reports in relation to this injury.
You may upload a scanned report/image or a smartphone picture of the report/image (*accepted file types include png, jpg, doc, xls, docx, xlsx, pdf, txt maximum file size 1 MB*).

Browse... Choose File No file selected Upload

130. If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?

- Yes
- No
- N/A

LOGIC Hidden unless: Question "If you do not have the medical reports, do you give us permission to contact your health care provider to obtain reports relevant to this injury?" #130 is one of the following answers ("Yes")

131. Please provide name and contact details of the medical provider who holds these records.

We will send you a permission slip to sign and a stamped envelope addressed to this medical provider.
This is necessary for the release of your records to us for this survey.

Name of medical provider:

Address of medical provider:

Phone contact for medical provider:

General Health Questions

LOGIC Show/hide trigger exists.

132. Have you ever smoked cigarettes?

- Yes
- No

LOGIC Dynamically shown if "Have you ever smoked cigarettes?" = Yes

133. If Yes, at what age did you start smoking?

LOGIC Show/hide trigger exists. Dynamically shown if "Have you ever smoked cigarettes?" = Yes

134. Do you currently smoke?

- Yes
- No

LOGIC Dynamically shown if "Do you currently smoke?" = Yes

135. If Yes, how many cigarettes per day on average?

LOGIC Dynamically shown if "Do you currently smoke?" = No

136. If No, at what age did you quit?

LOGIC Dynamically shown if "Do you currently smoke?" = No

137. When you were smoking, on average how many cigarettes per day would you smoke over the years until you quit?

LOGIC Show/hide trigger exists.

138. Do you ever consume alcoholic drinks?

- Yes
- No

LOGIC Dynamically shown if "Do you ever consume alcoholic drinks?" = Yes

139. If yes, approximately how many standard drinks would you consume per week? Examples of standard drinks include 100ml of wine, a bottle of mid-strength beer, a 285ml glass(midi) of full strength beer or a 30ml nip of spirits.

	1-3	4-6	7-10	10 +
drinks	1-3	4-6	7-10	10 +

General Health Questions

140. Have you ever been diagnosed with any of the following conditions/disorders?

	Yes	No
any type of cancer	○ Yes	◉ No
chronic renal failure	○ Yes	◉ No
rheumatoid arthritis	○ Yes	◉ No
osteoarthritis	○ Yes	◉ No
osteoporosis	○ Yes	◉ No
diabetes	○ Yes	◉ No
cystic fibrosis	○ Yes	◉ No
cerebral palsy	○ Yes	◉ No
cardiac conditions	○ Yes	◉ No
high blood pressure	○ Yes	◉ No
anaemia	○ Yes	◉ No
skin diseases	○ Yes	◉ No
thyroid disease	○ Yes	◉ No
gastrointestinal disease	○ Yes	◉ No
depression	○ Yes	◉ No
insomnia	○ Yes	◉ No
respiratory conditions	○ Yes	◉ No
neurological conditions	○ Yes	◉ No
other	○ Yes	◉ No

LOGIC Show/hide trigger exists.

141. Have you ever had hip, knee or ankle surgery?

- Yes
- No

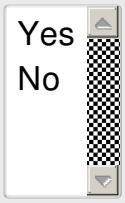
LOGIC Dynamically shown if "Have you ever had hip, knee or ankle surgery?" = Yes

142. If yes,

Hip

Knee

Ankle



LOGIC Show/hide trigger exists.

143. Have you ever had a fracture of any bone?

- Yes
- No

LOGIC Dynamically shown if "Have you ever had a fracture of any bone?" = Yes

144. If yes:

Where was the fracture? _____

When did it occur? _____

What treatment did you receive? _____

145. Do you have a family history of osteoporosis?

- Yes
- No
- Unsure

LOGIC Show/hide trigger exists.

146. To your knowledge, have you ever been treated using quinolone antibiotics (for example ciprofloxacin, norfloxacin)?

- Yes
- No
- Unsure

LOGIC Dynamically shown if "To your knowledge, have you ever been treated using quinolone antibiotics (for example ciprofloxacin, norfloxacin)?" = Yes

147. If yes, were you treated using these antibiotics within the 6 months prior to your injury?

- Yes
- No
- Unsure

148. To your knowledge, have you ever been treated using anti-seizure or epilepsy medications (for example clonazepam, gabapentin, lamotrigine, sodium valproate)?

- ○ Yes
- ○ No
- ○ Unsure

LOGIC Show/hide trigger exists.

149. To your knowledge, have you ever been treated using corticosteriod medication (for example cortisone injection, prednisone tablets, prednisolone tablets, flixotide inhaler, pulmicort inhaler, QVAR inhaler, seretide accuhaler, symbicort turbuhaler, steriod cream)?

- ○ Yes
- ○ No
- ○ Unsure

LOGIC Show/hide trigger exists. Dynamically shown if "To your knowledge, have you ever been treated using corticosteriod medication (for example cortisone injection, prednisone tablets, prednisolone tablets, flixotide inhaler, pulmicort inhaler, QVAR inhaler, seretide accuhaler, symbicort turbuhaler, steriod cream)?" = Yes

150. If yes, how was this drug administered?

- ☐ Injection
- ☐ Topical Cream
- ☐ Tablet
- ☐ Inhaler

LOGIC Dynamically shown if "If yes, how was this drug administered?" = Inhaler

151. Which type of inhaler?

General Health Questions

LOGIC Show/hide trigger exists.

152. To your knowledge, have you ever been treated using calcium tablets as prescribed by a medical doctor or taken it without a prescription?

- Yes
- No
- Unsure

LOGIC Show/hide trigger exists. Dynamically shown if "To your knowledge, have you ever been treated using calcium tablets as prescribed by a medical doctor or taken it without a prescription?" = Yes

153. Do you take calcium tablets on a regular basis?

- Yes
- No

LOGIC Dynamically shown if "Do you take calcium tablets on a regular basis?" = Yes

154. If yes, please provide details of dose and brand:

Dose

Brand

LOGIC Show/hide trigger exists.

155. To your knowledge, have you ever been treated using vitamin D supplementation as prescribed by a medical doctor or taken it without prescription?

- Yes
- No
- Unsure

LOGIC Show/hide trigger exists. Dynamically shown if "To your knowledge, have you ever been treated using vitamin D supplementation as prescribed by a medical doctor or taken it without prescription?" = Yes

156. Do you take vitamin D supplementation on a regular basis?

- Yes
- No

LOGIC Dynamically shown if "Do you take vitamin D supplementation on a regular basis?" = Yes

157. If yes, please provide details of dose and brand:

Dose

Brand

158. To your knowledge, have you ever

	Yes	No	Unsure
been treated using bisphosphonates (for example actonel, Didrocal, alendronate sodium, zoledronic acid)?	Yes	No	Unsure
undergone chemotherapy?	Yes	No	Unsure
undergone a bone marrow or organ transplant?	Yes	No	Unsure

LOGIC Show/hide trigger exists.

159. Are you regularly taking, or have you ever taken, any other regular medication as prescribed by a medical doctor?

- Yes
- No

LOGIC Dynamically shown if "Are you regularly taking, or have you ever taken, any other regular medication as prescribed by a medical doctor?" = Yes

160. If yes, please list medications

161. Are you regularly taking, or in the past have you regularly taken any other over-the-counter medications, dietary supplements or sports supplements?

- Yes
- No

LOGIC Hidden unless: Question "Are you regularly taking, or in the past have you regularly taken any other over-the-counter medications, dietary supplements or sports supplements?" #161 is one of the following answers ("Yes")

162. If yes, please list any other medications/supplements

- [ ] B-alanine
- [ ] B-Vitamins
- [ ] Creatine
- [ ] Fish Oil
- [ ] Glucosamine
- [ ] Iron
- [ ] Probiotics
- [ ] Protein Powder
- [ ] Vitamin C
- [ ] Zinc
- [ ] Other

Dietary Habits

163. Would you say that food dominates your life?

- ○ Yes
- ○ No

164. Do you:

	Yes	No
Currently suffer with or have you ever suffered in the past with an eating disorder?	Yes	No
Make yourself sick because you feel uncomfortably full?	Yes	No
Worry you have lost control over how much you eat?	Yes	No
Believe yourself to be fat when others say you are too thin?	Yes	No

165. Have you recently lost more than 6kgs in a 3 month period?

- Yes
- No

LOGIC Show/hide trigger exists.

166. Have you undergone any significant (greater than 5 kg) weight gain or loss in the past 2 years?

- Yes
- No

LOGIC Dynamically shown if "Have you undergone any significant (greater than 5 kg) weight gain or loss in the past 2 years?" = Yes

167. If Yes, please indicate the amount of weight gained/lost

	weight gain (kgs)
	weight lost (kgs)
	reason (if any known)

168. Do you follow any of the diets below:

	Yes	No
Vegetarian?	○ Yes	◉ No
Lacto-ovo vegetarian?	○ Yes	◉ No
Pesco-vegetarian?	○ Yes	◉ No
Vegan?	○ Yes	◉ No
Paleolithic/Paleo?	○ Yes	◉ No
Low carb/High Fat?	○ Yes	◉ No

LOGIC Show/hide trigger exists.

169. Do you follow a gluten free diet?

- ○ Yes
- ○ No

LOGIC Dynamically shown if "Do you follow a gluten free diet?" = Yes

170. If yes, please provide the reason why you follow this diet

- ○ Coeliac disease
- ○ Medically diagnosed gluten intolerance
- ○ Undiagnosed gluten intolerance
- ○ Irritable bowel syndrome
- ○ Other

LOGIC Show/hide trigger exists.

171. Do you have any food allergies or avoidances?

- ○ Yes
- ○ No

LOGIC Dynamically shown if "Do you have any food allergies or avoidances?" = Yes

172. If yes,

	Food type	Reason for avoidance
1	Food type	Reason for avoidance
2	Food type	Reason for avoidance
3	Food type	Reason for avoidance
4	Food type	Reason for avoidance

Female Health Questions

Page entry logic:
This page will show when: Question "Sex" is one of the following answers ("Female")

LOGIC Show/hide trigger exists.
173. Are you currently taking or have you ever taken the contraceptive pill?

- ○ Yes
- ○ No

LOGIC Dynamically shown if "Are you currently taking or have you ever taken the contraceptive pill?" = Yes
174. If yes, how long have you or did you take the contraceptive pill?

	< 1 year	1-2 years	3-5 years	5 + years
Length of time	○ < 1 year	○ 1-2 years	○ 3-5 years	○ 5 + years

LOGIC Dynamically shown if "Are you currently taking or have you ever taken the contraceptive pill?" = Yes
175. What was the name of the contraceptive pill you were taking?

LOGIC Show/hide trigger exists.
176. Do you have any children?

- ○ Yes
- ○ No

LOGIC Dynamically shown if "Do you have any children?" = Yes

177. Please indicate the year of birth for each child

Child 1

Child 2

Child 3

Child 4

Child 5

LOGIC Show/hide trigger exists.

178. What is the regular length of your menstrual cycle?

	I don't have a period	Irregular	< 26 days	27 - 31 days	> 31 days
Cycle length	I don't have a period	Irregular	< 26 days	27 - 31 days	> 31 days

LOGIC Dynamically shown if "What is the regular length of your menstrual cycle?" = I don't have a period

179. Please provide a reason why you don't have a period:

LOGIC Show/hide trigger exists.

180. Have you gone through menopause?

- Yes
- No

LOGIC Dynamically shown if "Have you gone through menopause?" = Yes

181. If yes, at what age did you go through menopause?

LOGIC Dynamically shown if "Have you gone through menopause?" = Yes

182. Was the menopause:

- Natural
- Surgical

183. Have you experienced amenorrhea (absence of menstrual periods) or oligomenorrhea (infrequent or irregular menstruation) or menorrhagia (excessive menstruation)?

- Yes
- No

184. What age did you commence your menstrual period?

	Early than 12 years old	12 years old	13 years old	14 years old	15 years old	Later than 15 years old
Age	Early than 12 years old	12 years old	13 years old	14 years old	15 years old	Later than 15 years old

Complete

Thank You for Participating!

If you have indicated and are eligible, you may be contacted in the future to provide a saliva sample for analysis.

You have been entered into the draw to win one of 50 vouchers to spend $50 high performance sportswear from 2XU.COM.AU.

For more information on the Collaborative Research Network's research activities please visit the CRN website.

If you have any further questions, please do not hesitate to contact Dr Nicole Vlahovich or Dr Renae Domaschenz phone (02) 6214 7319 or email injurystudy@ausport.gov.au

See terms and conditions for voucher giveaway.
Permit No ACT TP 14/03056

Appendix 6 – 'AIS Running Injury Study' Communication plan

Injury Study Recruitment Plan

Online Survey

Participants will be recruited to complete an online survey using the following methods:

- Via medical clinics: Physical Therapy, Radiology and Sports Medicine clinics will be contacted and asked to be involved in the recruitment of participants to the online survey. After making initial contact with the head office/CEO/CMO, the clinic will be provided with flyers directing potential participants to the URL of the online survey. Initial contact will be made via letter, email or in person by Dr David Hughes. A master list has been created of the major clinics in Australia, their contact details and the person who would be responsible for 'gatekeeper approval' for the promotion of the study within the clinic. It is also intended that Dr Hughes will utilise an extended network of medical providers. It is anticipated that these clinics will be critical in the recruitment of participants in the 'injured' cohorts.
- Via running races/ events: Race and event organisers will be contacted initially via email or letter asking to promote the online survey to people who are registering online for the event. Following online registration to an event it is typical that the entrant receives a confirmation email, we would ask that the event organisers include a link to the URL of the online survey in the confirmation email to participants. A master list has been created of the major running events in Australia, the contact details of the organisation holding these events and the person who would be responsible for 'gatekeeper approval' for the promotion of the study in association with the event.
- Advertisements in running magazines and websites: Advertisements will be placed in running related publications leading potential participants to the URL of the online survey.
- Social Media: Social media accounts will be set up for the CRN injury study including Twitter, Facebook and Google + to promote the study and participation to a wider, more general audience. These channels may also be used to cross promote other studies within the CRN and release interesting information, research updates, media articles that are related to the study.
- Media: Opportunities may arise where Dr Hughes is able to promote the study via the media, this will be done on an ad hoc basis.

Each of the approaches will provide a direct link to the URL of the online survey.

As the inclusion criteria to participate in the online survey is quite broad it is expected that broad based promotion, such as through social media, event websites and running publications, will be quite effective.

The point of contact for the participants will be a common email address and phone number that will be staffed by Nicole Vlahovich, Renae Domaschenz or the project's PhD student. These people will also be responsible for updating, and responding to queries through, social media channels.

Copies of template emails/letters, advertising material will be approved by the ethics committee.

Genetic testing

Participants for genetic testing will be recruited directly from the survey participants. Survey responses to particular questions, including injury state and ethnicity, will be sorted to develop a pool of potential participants. Those meeting the criteria for inclusion into the genetics study will be contacted by phone or email in order to explain the project and extend an invitation to participate.

A participation pack, including the saliva collection device, and the participant information sheets and consent forms, will then be mailed out to the participant.

This part of the study will not be promoted through media or other channels as the online survey acts as the recruitment tool for the participation in the genetic study.

Appendix 7 – Facebook advertisement example

Australian Institute of Sport
Written by Digi Comms [?] · 1 min ·

Runners who complete the survey receive a 20 per cent discount offer on 2XU compression

Join running study

Runners who complete the survey receive a 20 per cent discount offer on 2XU compression from 2XU.COM.AU (offer ends 31 May 2016 and terms and conditions apply).

WWW.AUSPORT.GOV.AU

Learn More

Boost Post

Like Comment Share

Write a comment...

Appendix 8 – Participant information about the genetic research

Participant Information Sheet/Consent Form

A study to investigate the relationship between genes, physical activity and health status.

Title	*The genetics of exercise-induced injuries involving tendon and bone*
Protocol Number	*RO21688B*
Project Sponsor	*CRN for Advancing Exercise & Sports Science*
Principal Investigator	*Dr David Hughes*
Associate Investigator(s)	*Dr Nicole Vlahovich, Maria Kozlovskaia (Department of Sports Medicine, Australian Institute of Sport), A/Prof Bon Gray, A/Prof Lotti Tajouri, A/Prof Justin Keogh, A/Prof Mike Climstein, Rebecca Grealy, A/Prof Kevin Ashton and Professor Nuala Byrne (Bond University), Professor Matthew Brown and Dr Paul Leo (University of Queensland) and Professor Maria Fiatarone Singh, Dr Yorgi Mavros, Guy Wilson and Jacinda Meiklejohn (University of Sydney).*
Location	*Australian Institute of Sport, Bond University & CRN partner organisations*

Introduction

You are invited to take part in this research project because you completed the online survey for the study *the genetics of exercise-induced injuries involving tendon and bone*, agreeing to follow up genetic analysis and you are:

- A recreational runner,
- Aged 18-50 years of age,
- Run greater than 15km per week, and
- Fall into one of the following categories:
    - Uninjured* for the past two years (*no injury below the knee due to running that required you to discontinue running for a period of 2 weeks or more)
    - Diagnosed with a below knee bone stress injury caused by running, by a physician or physiotherapist and confirmed by diagnostic imaging, in the last 2 years
    - Diagnosed with Achilles tendinopathy or Achilles tendinitis caused by running, by a physician or physiotherapist in the last 2 years

The Participant Information and Consent Forms tell you about the research project. It explains the procedures involved. Knowing what is involved will help you decide if you want to take part in the research.

Participant Number: AIS

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not to take part, you might want to talk about it with a relative, friend or healthcare worker.

Participation in this research is voluntary. If you don't wish to take part, you don't have to. You will receive the best possible care whether you take part or not. You're not obliged to participate and if you do, you can withdraw at any time without penalty or prejudice and any samples that you have provided will be disposed of.

If you agree to participate in this study, we would like you to complete the enclosed paperwork and provide a saliva sample according to the enclosed instructions.

All aspects of the study including your personal details and results will be strictly confidential and only the principal researchers above will have access to this information.

If you decide you want to take part in the research project, you will be asked to sign the consent section. By signing it you are telling us that you:

- understand what you have read;
- consent to take part in the research project;
- consent to participate in the research processes that are described;
- consent to the use of your personal and health information as described
- certify to the best of your knowledge and belief, you have no physical or mental illness or weakness that would increase the risk to of participating in this investigation;
- are participating in this project of your own free will and have not been coerced in any way to participate;

You will be given a copy of this Participant Information and Consent Form to keep.

What is the purpose of this study?
Identifying sequences or new mutations in known genes will help researchers to better understand the relationship between lifestyle, health status and genetic profile and we hope that this information will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases. The genetic sequence information will provide new insights into the genetic factors associated with exercise-induced injuries in recreational and elite athletes.

This research is being conducted by a collaboration between Bond University, the Australian Institute of Sport, the University of Queensland and the University of Sydney.

What does participation in the research project involve?
Participation is this portion of the project involves the provision of a saliva sample for genetic analysis. Enclosed are the consent forms to read and sign, a kit that will help you collect the specimen and return envelope to return your sample. This process should take approximately 10-15 minutes.

What are the possible benefits to participating?
We cannot guarantee that you will receive any benefits from this research, but your participation in the study may help doctors to better understand the relationship between lifestyle, health status and genetic profile with the hope that this will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases.

What will happen to my test sample and what are the possible risks to participating in the study?
Please see the document *The Ethics of Genetic Research*

What if new information arises?
During the research project, new information about the risks and benefits of the project may become known to the researchers. If this occurs, you will be told about this new information and the researcher will discuss whether this new information affects you.

Can anyone participate in this study?
As long as you meet the criteria specified earlier, and have participated in the online survey, you are eligible to take part.

Do I have to take part in this research project?
If you do not wish to take part then you don't have to. If you decide to take part and later change your mind, you are free to withdraw at any stage. All information that you have provided can be destroyed at any time. You can withdraw your consent to participate in this research project by emailing the Principal Investigator at injurystudy@ausport.gov.au

Before you make a decision to participate in any follow up studies, a member of the research team will contact you so that you can ask any questions you have about the project. You can ask for any information that you want.

Is this research project approved?

This project will be carried out according to the *National Statement on Ethical Conduct in Research Involving Humans* (2007) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies.

The ethical aspects of this research project have been approved by Bond University Human Research Ethics Committee; protocol number RO21688B, contact Dr Lisa Marlow, Research Ethics Manager on (07) 5595 4194.

Will I get paid to participate in this study?
You will not be paid for participating in this study but all costs such as expenses involved with any investigation, posting of forms, questionnaires and samples will be covered

Who can I contact if I have any questions or problems in relation to this study?
If you wish to discuss further the experimental procedure or have any questions, please do not hesitate to contact Dr Nicole Vlahovich phone (02) 6214 7319 or email injurystudy@ausport.gov.au
If you have any concerns with respect to the conduct of this study, you may contact the Secretary of the Bond University Human Research Ethics Committee Dr Lisa Marlow on (07) 5595 4194 or by email lmarlow@bond.edu.au

Participant Number: AIS

The genetics of exercise-induced injuries involving tendon and bone **information sheet**

Genetic Sample and Data Consent

Principal Investigator: Dr David Hughes

Statement of Informed Consent for Genetic Sample and Data Collection

I have read, or had read to me in a language that I understand, this Participant Information Sheet and I understand the purposes, procedures and risks of this research project as described within it. I have had the opportunity to ask questions and I am satisfied with the answers I have received.

☐ *I give permission for my anonymised sample and/or clinical information to be shared by the Investigators of this study with collaborating researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

☐ *I give permission for my anonymised sample and/or clinical information to be part of the Biobank, to be shared with other researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

I understand that there is a very small chance that this study could identify a genetic defect that increases my risk for an unrelated condition.

☐ I **would like to be informed** if a risk factor for a treatable condition is identified

☐ I **would NOT like to be informed** if a risk factor for a treatable condition is identified

I understand that I will not receive genetic results for conditions where there is no known treatment at the time the result becomes available.

I understand that, due to the type of genetic test being performed, participation in this research project will not provide me with a clearance from any genetic or heritable conditions.

I freely agree to participate in this research project as described.

Name of Participant	Date	Signature of Participant

Name of Researcher	Date	Signature of Researcher

Appendix 9 – The ethics of genetic research

The Ethics and Risks Associated with Genetic Research

COLLABORATIVE RESEARCH NETWORK
ADVANCING EXERCISE & SPORTS SCIENCE

A study to investigate the relationship between genes, physical activity and health status.

What is genetic research?
Genes are made of DNA – the chemical structure carrying your genetic information that determines many human characteristics such as the colour of your eyes. Researchers study how genes are expressed in order to understand why some people differ in their response to exercise and to develop, improved and individualised approaches to exercise prescription. This will potentially provide new ways in which to reduce disease within the wider community

What will happen to my test sample?
Saliva samples will be used immediately or stored for future studies. Saliva samples will be used to extract your DNA. DNA sequence and other information obtained in this research project will be stored securely, with each participant sample allocated a specific code number. All data processing will occur by use of this number and not the participant's name. However all samples are potentially re-identifiable and must remain so to satisfy the aims of the research and also to identify participants if required for the return of genetic information.

The samples will be tested using genome-wide association studies. These studies are increasingly being used to identify biological pathways and networks underlying complex diseases. This type of genetic test is very unlikely to identify mutations associated with heritable diseases as it will not provide information about the whole genome, only portions of the genome which are unlikely to contain genes relating to disease.

Researchers from the Australian Institute of Sport (Canberra), University of Queensland (Brisbane), the University of Sydney and Bond University (Gold Coast, Queensland) are collaborating to study your DNA samples for research only. We may also share some information about your clinical information in our research group but all identifiers will be removed. All medical information is stored in password protected databases.

We will also be establishing a biobank – this is a way of collecting and storing anonymous biological material and information for further research. If you consent, your DNA will be stored and then can be used by other researchers for future projects.

What are the possible risks of participating in the study?
All medical information is stored in password protected databases. It is possible, though highly unlikely, that someone could get access to this database without permission. All samples and participant data will be stored securely in locked facilities.

This type of genetic test is very unlikely to identify mutations associated with heritable diseases. However, as part of this study there is a small chance that we might coincidentally find a defect in a gene which, in the duration of the project, may be identified as being associated with an increased risk for a genetic condition. If the condition is treatable or preventable, you can specify on the consent form if you want to be informed about such a finding. Learning this information may be upsetting. It could also affect your ability to get life insurance. The information could be used against you in the work setting.

You may be asked to give us health information about your relatives. Any information you give us will be kept confidential. We will not contact your relatives without your permission. We may discuss with you the possibility of including your relatives in the research project in the future. You may learn information from your test result about inherited diseases or disorders that may affect others, such as your brothers or sisters. This could interfere with family relationships. You may be faced with the decision to make the family aware of the existence of genetic information. Family members may or may not wish to know this information. Therefore, although we think that there are benefits in having that information, individuals have to weigh up the risks and decide whether or not to receive that information. In the unlikely event that a risk for an unrelated genetic condition is identified we will help to put you in contact with a local genetics service.

If we identify that you could be at risk for a disorder, which is currently untreatable and unpreventable, we will not disclose this information.

If you become upset or distressed as a result of your participation in the research, the researcher is able to arrange for counselling or other appropriate support. Any counselling or support will be provided by staff, who are not members of the research team. In addition, you may prefer to suspend or end your participation in the research if distress occurs.

What if new information arises?
During the research project, new information about the risks and benefits of the project may become known to the researchers. If this occurs, you will be told about this new information and the researcher will discuss whether this new information affects you.

How will I be informed of results from of this research project?
In relation to any genetic information generated from use of your DNA, you may elect to:

1. Have no information returned to you.
2. If a risk factor for a treatable condition is identified, have information returned to you to provide to your medical practitioner for consideration, verification testing and possible clinical or other action.

If your preference is not to receive genetic information, a researcher will contact you following the completion of genetic testing to ensure that this preference still remains.

What will happen to information about me?

In accordance with relevant Australian privacy and other relevant laws, you have the right to access the information collected and stored by the researchers about you. You also have the right to request that any information, with which you disagree, be corrected. Please contact one of the researchers named at the end of this document if you would like to access your information.

If you wish to discuss further the experimental procedure or have any questions, please do not hesitate to contact Dr Nicole Vlahovich phone (02) 6214 7319 or email injurystudy@ausport.gov.au

If you have any concerns with respect to the conduct of this study, you may contact the Secretary of the Bond University Human Research Ethics Committee Dr Lisa Marlow on (07) 5595 4194 or by email lmarlow@bond.edu.au

Appendix 10 – Mail package with consent forms

COLLABORATIVE RESEARCH NETWORK
ADVANCING EXERCISE & SPORTS SCIENCE

Thank you for agreeing to participate in *The genetics of exercise induced injuries involving tendon and bone.*

Please set aside 5 minutes to collect the sample and fill out the paperwork.

Please follow the steps below, **skipping a step could mean that your sample is unable to be included in the research**:

1. You must **wait 30 minutes after eating, drinking or smoking** before providing a saliva sample.
2. Read the Participant Information Sheet and the document entitled *The Ethics of Genetic Research - ethically defensible plan.* Contact the researchers using the email/phone number provided if you have any questions.
3. Once the information is read and understood, sign the consent form (the form marked COPY is for you to keep for your own reference).
4. Open the Oragene – DNA kit and carefully read the instructions provided on the next page, follow the five steps carefully. Please ensure that the sample fills the tube to the line without any bubbles
5. Place the **collection tube** (but not the funnel) **into the biohazard bag**.
6. Place the following in the reply paid envelope:
    a. **Tube containing saliva sample in the biohazard bag**
    b. **The signed consent form (single page only)**

    then seal and place the envelope in the post.

Please do not hesitate to contact the AIS if you have any questions or issues.

Kind regards,

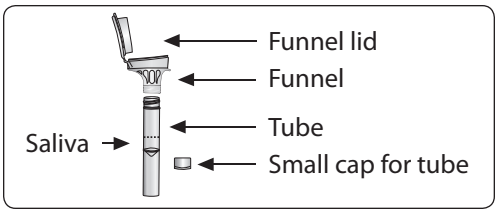
Nicole Vlahovich
02 6214 7319 or 0406376633
injurystudy@ausport.gov.au

ORAgene®•DNA

REF OG-500/510

USER INSTRUCTIONS

Read all instructions prior to collection



Collection precautions:

Do NOT eat, drink, smoke or chew gum for 30 minutes before giving your saliva sample.

Do NOT remove the plastic film from the funnel lid.

Intended use: For the collection of human DNA from saliva samples.

Contents: Kit contains stabilizing liquid.

Warnings and precautions: Wash with water if stabilizing liquid comes in contact with eyes or skin. Do NOT ingest. See MSDS at www.dnagenotek.com.

Small cap, choking hazard.

Storage: 15°C–30°C

Summary and explanation of the kit:

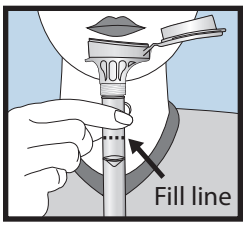
Oragene•DNA is a self-collection kit that provides the materials and instructions for collecting and stabilizing saliva specimens.

Label legend:

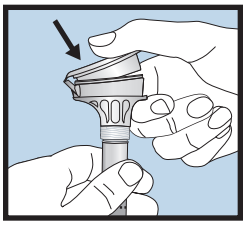
	Consult package insert
	Collect saliva by (Use by)
IVD	In vitro diagnostic medical device
REF	Catalog number
CE	CE Marking
	Caution, consult instructions for use
15°C–30°C	Storage instructions
EC REP	Authorized Representative
	Manufacturer

Procedure:

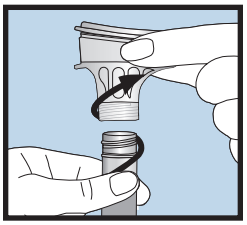
Most people take between 2 and 5 minutes to deliver a saliva sample following steps 1 to 5.



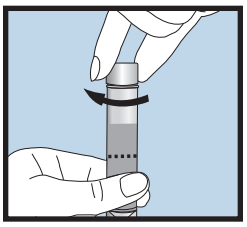
1 Spit into funnel until the amount of liquid saliva (not bubbles) reaches the fill line shown in picture #1.



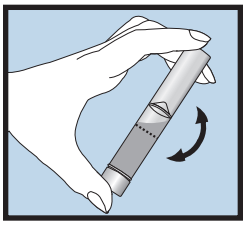
2 Hold the tube upright with one hand. Close the funnel lid with the other hand (as shown) by firmly pushing the lid until you hear a loud click. The liquid in the lid will be released into the tube to mix with the saliva. Make sure that the lid is closed tightly.



3 Hold the tube upright. Unscrew the funnel from the tube.



4 Use the small cap to close the tube tightly.



5 Shake the capped tube for 5 seconds. Discard or recycle the funnel.

CE IVD

For In Vitro Diagnostic Use

DNA genotek

Superior samples
Proven performance

Made in Canada
DNA Genotek Inc.
Ottawa, ON, Canada K2K 1L1

Tel.: 613.723.5757
info@dnagenotek.com
www.dnagenotek.com

EC REP Emergo Europe, Molenstraat 15, 2513 BH The Hague, The Netherlands
Tel: (+31) (0) 70 345-8570 Fax: (+31) (0) 70 346-7299

Australian Sponsor: Emergo Australia, Level 20, Tower II, Darling Park, 201 Sussex Street, Sydney, NSW 2000 Australia

PD-PR-118 Issue 3/2014-06

Participant Information Sheet/Consent Form

A study to investigate the relationship between genes, physical activity and health status.

Title	*The genetics of exercise-induced injuries involving tendon and bone*
Protocol Number	*RO21688B*
Project Sponsor	*CRN for Advancing Exercise & Sports Science*
Principal Investigator	*Dr David Hughes*
Associate Investigator(s)	*Dr Nicole Vlahovich, Maria Kozlovskaia (Department of Sports Medicine, Australian Institute of Sport), A/Prof Bon Gray, A/Prof Lotti Tajouri, A/Prof Justin Keogh, A/Prof Mike Climstein, Rebecca Grealy, A/Prof Kevin Ashton and Professor Nuala Byrne (Bond University), Professor Matthew Brown and Dr Paul Leo (University of Queensland) and Professor Maria Fiatarone Singh, Dr Yorgi Mavros, Guy Wilson and Jacinda Meiklejohn (University of Sydney).*
Location	*Australian Institute of Sport, Bond University & CRN partner organisations*

Introduction

You are invited to take part in this research project because you completed the online survey for the study *the genetics of exercise-induced injuries involving tendon and bone*, agreeing to follow up genetic analysis and you are:

- A recreational runner,
- Aged 18-50 years of age,
- Run greater than 15km per week, and
- Fall into one of the following categories:
    - Uninjured* for the past two years (*no injury below the knee due to running that required you to discontinue running for a period of 2 weeks or more)
    - Diagnosed with a below knee bone stress injury caused by running, by a physician or physiotherapist and confirmed by diagnostic imaging, in the last 2 years
    - Diagnosed with Achilles tendinopathy or Achilles tendinitis caused by running, by a physician or physiotherapist in the last 2 years

The Participant Information and Consent Forms tell you about the research project. It explains the procedures involved. Knowing what is involved will help you decide if you want to take part in the research.

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether or not to take part, you might want to talk about it with a relative, friend or healthcare worker.

Participation in this research is voluntary. If you don't wish to take part, you don't have to. You will receive the best possible care whether you take part or not. You're not obliged to participate and if you do, you can withdraw at any time without penalty or prejudice and any samples that you have provided will be disposed of.

If you agree to participate in this study, we would like you to complete the enclosed paperwork and provide a saliva sample according to the enclosed instructions.

All aspects of the study including your personal details and results will be strictly confidential and only the principal researchers above will have access to this information.

If you decide you want to take part in the research project, you will be asked to sign the consent section. By signing it you are telling us that you:

- understand what you have read;
- consent to take part in the research project;
- consent to participate in the research processes that are described;
- consent to the use of your personal and health information as described
- certify to the best of your knowledge and belief, you have no physical or mental illness or weakness that would increase the risk to of participating in this investigation;
- are participating in this project of your own free will and have not been coerced in any way to participate;

You will be given a copy of this Participant Information and Consent Form to keep.

What is the purpose of this study?
Identifying sequences or new mutations in known genes will help researchers to better understand the relationship between lifestyle, health status and genetic profile and we hope that this information will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases. The genetic sequence information will provide new insights into the genetic factors associated with exercise-induced injuries in recreational and elite athletes.

This research is being conducted by a collaboration between Bond University, the Australian Institute of Sport, the University of Queensland and the University of Sydney.

What does participation in the research project involve?
Participation is this portion of the project involves the provision of a saliva sample for genetic analysis. Enclosed are the consent forms to read and sign, a kit that will help you collect the specimen and return envelope to return your sample. This process should take approximately 10-15 minutes.

What are the possible benefits to participating?
We cannot guarantee that you will receive any benefits from this research, but your participation in the study may help doctors to better understand the relationship between lifestyle, health status and genetic profile with the hope that this will ultimately improve health and quality of life, prevent injury or disability, and prevent or treat chronic diseases.

What will happen to my test sample and what are the possible risks to participating in the study?
Please see the document *The Ethics of Genetic Research*

What if new information arises?
During the research project, new information about the risks and benefits of the project may become known to the researchers. If this occurs, you will be told about this new information and the researcher will discuss whether this new information affects you.

Can anyone participate in this study?
As long as you meet the criteria specified earlier, and have participated in the online survey, you are eligible to take part.

Do I have to take part in this research project?
If you do not wish to take part then you don't have to. If you decide to take part and later change your mind, you are free to withdraw at any stage. All information that you have provided can be destroyed at any time. You can withdraw your consent to participate in this research project by emailing the Principal Investigator at injurystudy@ausport.gov.au

Before you make a decision to participate in any follow up studies, a member of the research team will contact you so that you can ask any questions you have about the project. You can ask for any information that you want.

Is this research project approved?

This project will be carried out according to the *National Statement on Ethical Conduct in Research Involving Humans* (2007) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies.

The ethical aspects of this research project have been approved by Bond University Human Research Ethics Committee; protocol number RO21688B, contact Dr Lisa Marlow, Research Ethics Manager on (07) 5595 4194.

Will I get paid to participate in this study?
You will not be paid for participating in this study but all costs such as expenses involved with any investigation, posting of forms, questionnaires and samples will be covered

Who can I contact if I have any questions or problems in relation to this study?
If you wish to discuss further the experimental procedure or have any questions, please do not hesitate to contact Dr Nicole Vlahovich phone (02) 6214 7319 or email injurystudy@ausport.gov.au
If you have any concerns with respect to the conduct of this study, you may contact the Secretary of the Bond University Human Research Ethics Committee Dr Lisa Marlow on (07) 5595 4194 or by email lmarlow@bond.edu.au

***The genetics of exercise-induced injuries involving tendon and bone* information sheet**

Genetic Sample and Data Consent

Principal Investigator: Dr David Hughes

Statement of Informed Consent for Genetic Sample and Data Collection

I have read, or had read to me in a language that I understand, this Participant Information Sheet and I understand the purposes, procedures and risks of this research project as described within it. I have had the opportunity to ask questions and I am satisfied with the answers I have received.

☐ *I give permission for my anonymised sample and/or clinical information to be shared by the Investigators of this study with collaborating researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

☐ *I give permission for my anonymised sample and/or clinical information to be part of the Biobank, to be shared with other researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

I understand that there is a very small chance that this study could identify a genetic defect that increases my risk for an unrelated condition.

☐ I **would like to be informed** if a risk factor for a treatable condition is identified

☐ I **would NOT like to be informed** if a risk factor for a treatable condition is identified

I understand that I will not receive genetic results for conditions where there is no known treatment at the time the result becomes available.

I understand that, due to the type of genetic test being performed, participation in this research project will not provide me with a clearance from any genetic or heritable conditions.

I freely agree to participate in this research project as described.

Name of Participant	Date	Signature of Participant

Name of Researcher	Date	Signature of Researcher

The Ethics and Risks Associated with Genetic Research

COLLABORATIVE RESEARCH NETWORK
ADVANCING EXERCISE & SPORTS SCIENCE

A study to investigate the relationship between genes, physical activity and health status.

What is genetic research?
Genes are made of DNA – the chemical structure carrying your genetic information that determines many human characteristics such as the colour of your eyes. Researchers study how genes are expressed in order to understand why some people differ in their response to exercise and to develop, improved and individualised approaches to exercise prescription. This will potentially provide new ways in which to reduce disease within the wider community

What will happen to my test sample?
Saliva samples will be used immediately or stored for future studies. Saliva samples will be used to extract your DNA. DNA sequence and other information obtained in this research project will be stored securely, with each participant sample allocated a specific code number. All data processing will occur by use of this number and not the participant's name. However all samples are potentially re-identifiable and must remain so to satisfy the aims of the research and also to identify participants if required for the return of genetic information.

The samples will be tested using genome-wide association studies. These studies are increasingly being used to identify biological pathways and networks underlying complex diseases. This type of genetic test is very unlikely to identify mutations associated with heritable diseases as it will not provide information about the whole genome, only portions of the genome which are unlikely to contain genes relating to disease.

Researchers from the Australian Institute of Sport (Canberra), University of Queensland (Brisbane), the University of Sydney and Bond University (Gold Coast, Queensland) are collaborating to study your DNA samples for research only. We may also share some information about your clinical information in our research group but all identifiers will be removed. All medical information is stored in password protected databases.

We will also be establishing a biobank – this is a way of collecting and storing anonymous biological material and information for further research. If you consent, your DNA will be stored and then can be used by other researchers for future projects.

What are the possible risks of participating in the study?
All medical information is stored in password protected databases. It is possible, though highly unlikely, that someone could get access to this database without permission. All samples and participant data will be stored securely in locked facilities.

This type of genetic test is very unlikely to identify mutations associated with heritable diseases. However, as part of this study there is a small chance that we might coincidentally find a defect in a gene which, in the duration of the project, may be identified as being associated with an increased risk for a genetic condition. If the condition is treatable or preventable, you can specify on the consent form if you want to be informed about such a finding. Learning this information may be upsetting. It could also affect your ability to get life insurance. The information could be used against you in the work setting.

You may be asked to give us health information about your relatives. Any information you give us will be kept confidential. We will not contact your relatives without your permission. We may discuss with you the possibility of including your relatives in the research project in the future. You may learn information from your test result about inherited diseases or disorders that may affect others, such as your brothers or sisters. This could interfere with family relationships. You may be faced with the decision to make the family aware of the existence of genetic information. Family members may or may not wish to know this information. Therefore, although we think that there are benefits in having that information, individuals have to weigh up the risks and decide whether or not to receive that information. In the unlikely event that a risk for an unrelated genetic condition is identified we will help to put you in contact with a local genetics service.

If we identify that you could be at risk for a disorder, which is currently untreatable and unpreventable, we will not disclose this information.

If you become upset or distressed as a result of your participation in the research, the researcher is able to arrange for counselling or other appropriate support. Any counselling or support will be provided by staff, who are not members of the research team. In addition, you may prefer to suspend or end your participation in the research if distress occurs.

What if new information arises?
During the research project, new information about the risks and benefits of the project may become known to the researchers. If this occurs, you will be told about this new information and the researcher will discuss whether this new information affects you.

How will I be informed of results from of this research project?
In relation to any genetic information generated from use of your DNA, you may elect to:

1. Have no information returned to you.
2. If a risk factor for a treatable condition is identified, have information returned to you to provide to your medical practitioner for consideration, verification testing and possible clinical or other action.

If your preference is not to receive genetic information, a researcher will contact you following the completion of genetic testing to ensure that this preference still remains.

What will happen to information about me?

In accordance with relevant Australian privacy and other relevant laws, you have the right to access the information collected and stored by the researchers about you. You also have the right to request that any information, with which you disagree, be corrected. Please contact one of the researchers named at the end of this document if you would like to access your information.

If you wish to discuss further the experimental procedure or have any questions, please do not hesitate to contact Dr Nicole Vlahovich phone (02) 6214 7319 or email injurystudy@ausport.gov.au

If you have any concerns with respect to the conduct of this study, you may contact the Secretary of the Bond University Human Research Ethics Committee Dr Lisa Marlow on (07) 5595 4194 or by email lmarlow@bond.edu.au

***The genetics of exercise-induced injuries involving tendon and bone* information sheet**

Genetic Sample and Data Consent

Principal Investigator: Dr David Hughes

Statement of Informed Consent for Genetic Sample and Data Collection

I have read, or had read to me in a language that I understand, this Participant Information Sheet and I understand the purposes, procedures and risks of this research project as described within it. I have had the opportunity to ask questions and I am satisfied with the answers I have received.

☐ *I give permission for my anonymised sample and/or clinical information to be shared by the Investigators of this study with collaborating researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

☐ *I give permission for my anonymised sample and/or clinical information to be part of the Biobank, to be shared with other researchers who have ethically approved studies and are researching the relationship between genes, physical activity and health status.*

I understand that there is a very small chance that this study could identify a genetic defect that increases my risk for an unrelated condition.

☐ I **would like to be informed** if a risk factor for a treatable condition is identified

☐ I **would NOT like to be informed** if a risk factor for a treatable condition is identified

I understand that I will not receive genetic results for conditions where there is no known treatment at the time the result becomes available.

I understand that, due to the type of genetic test being performed, participation in this research project will not provide me with a clearance from any genetic or heritable conditions.

I freely agree to participate in this research project as described.

Participant COPY

Name of Participant	Date	Signature of Participant
Name of Researcher	Date	Signature of Researcher